CasProt

D1.1 Report on protein stability and folding

Name of project: Fostering high scientific quality in protein research in Eastern Slovakia (**CasProt**)

No. 952333

D1.1 Report on protein stability and folding

Name of project: Fostering high scientific quality in protein research in Eastern Slovakia (**CasProt**)

No. 952333

Deliverable No:	D.1.1.
Work package no. and title	WP1 Research activities and know-how transfer
Task no. and title:	Task 1.3. Report on protein stability and folding
Nature:	ORDP : Open Research Data Pilot
Dissemination level:	Public
Lead Beneficiary:	UPJS
Author(s):	Assoc. prof. RNDr. Gabriel Žoldák, PhD. RNDr. Veronika Džupponová MSc. Michal Gala
Reviewer:	Assoc. prof. Daniel Jancura, PhD. Abhigyan Sengupta, PhD. Prof. Matthias Rief
Date of publication:	30th of March 2022

Document History

Date	Authors	Version
30th of March 2022	Assoc. prof. RNDr. Gabriel Žoldák, PhD., RNDr. Veronika Džupponová,MSc. Michal Gala	1.0

APPROVALS

Task leader: Assoc. prof. RNDr. Gabriel Žoldák, PhD.
WP leader: Assoc. prof. RNDr. Erik Sedlák, PhD.
Project coordinator: Assoc. prof. RNDr. Erik Sedlák, PhD.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

ABBREVIATIONS

°C	celsius
µM	micromolar
2-PMT	2-phenylmelatonin
3D	three-dimensional
Å	ångström
AAB	collagen type 1, A chain – A chain – B chain
AC/DC	acrylamide carrier/dopant copolymer
AFM	atomic force microscope
AIMP2-	DX2 AIMP2 lacking exon 2
AL	amyloid light-chain
Ala	alanine
AMPs	antimicrobial peptides
AOs	amyloid oligomers
AR	androgen receptor
Arg	arginine
ArM	artificial metalloenzyme
ATP	adenosine triphosphate
BiP	immunoglobulin binding protein
CETSA	cellular thermal shift assay
CFTR	cystic fibrosis transmembrane conductance regulator
CI2	chymotrypsin inhibitor 2
C-IDR	C-terminal intrinsically disordered region
Cnx	calnexin
cryo-EM	cryoelectron microscopy
CS	chemical shifts
CsgA	major curlin subunit
Cx32	connexin 32
Cys	cysteine
D2	receptor dopamine receptor D2
DBD	DNA binding domain
DDR	discoidin domain receptor
DEER	double electron electron resonance
DNA	deoxyribonucleic acid
DoxDHCer	deoxydihydroceramide
EGF	epidermal growth factor
eIF2α	eukaryotic initiation factor 2α
EMC	ER-membrane complex
ER	endoplasmic reticulum
ERAD	endoplasmic-reticulum-associated protein degradation
FABP	fatty acid-binding protein
FJC	freely jointed chain

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

FLNC	filamin C
FMN	flavin mononucleotide
GDP	guanosine diphosphate
GEM	genome-scale metabolic models
Glu	glutamic acid
GLP-1	glucagon-like peptide1
Gly	glycine
GOF	gain of function
GPCR	G protein-coupled receptor
GPR52	G Protein-Coupled Receptor 52
GUN	genome uncoupled
H/D	hydrogen-deuterium
H8	C-terminal amphipathic helix 8
HA	hemagglutinin
HBP	histidine-rich squid beak proteins
HCV	gepatitis C virus
HDX-MS	hydrogen-deuterium exchange mass spectrometry
hnRNPA1	heterogeneous nuclear ribonucleoprotein A1
HRI	Heme-regulated eIF2α kinase
HSF1	heat-shock factor 1
HSP70	heat shock protein 70
HSP90	heat shock protein 90
HspB1	heat shock protein family B member 1
HSPB8	heat shock protein family B member 8
IAP	inhibitor of apoptosis protein
IAPP	islet amyloid polypeptide
IDR	intrinsically disordered region
IgA	immunoglobulin A
IL	interleukin
IRE1	inositol-requiring enzyme 1
JDPs	J-domain proteins
Kapβ2	karyopherin-β2
LacY	single lactose permease
LC	light-chain
LLPS	liquid-liquid phase separation
LOF	lose of function
LUCS	loop-helix-loop unit combinatorial sampling
MaSIF	molecular surface interaction fingerprinting
MD	molecular dynamics
Met	methionine
MG132	carbobenzoxy-Leu-Leu-leucinal
MKP3	dual-specificity phosphatase 6
mM	millimolar
mRNA	messenger RNA
MS	mass spectrometry
MT2	melatonin 2
mtHSP70	mitochondrial heat shock protein 70
NBD1	first nucleotide-binding domain
nm	nanometer
NMR	nuclear magnetic resonance

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

NTD	N-terminal domain
OptoMB	light-controlled monobody
PAG1	proteasome α subunit G1
PDI	protein disulfide isomerase
PERK	protein kinase R-like endoplasmic reticulum kinase
PIN	proteome integrity number
pN	piconewton
PPI	protein-protein interaction
Pro	proline
RBD	receptor-binding domain
Ric8A	resistance to inhibitors of cholinesterase-8A
ROC	receiver operating characteristic
RQC	ribosome-associated protein quality control
RTK	receptor tyrosine kinase
s	second
SA	small ribosomal
SAA1	serum amyloid A1
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SBMA	spinobulbar muscular atrophy
SCNG	single-chain nanogel
SE	serrate
SEC-MALS	size-exclusion chromatography coupled to multi-angle light scattering
Ser	serine
SMFS	single-molecule force spectroscopy
SSNMR	solid-state NMR
STIM1	stromal Interaction Molecule 1
TCR	T-cell receptor
TDP-43	TAR DNA-binding protein 43
TF	trigger factor
ThT	thioflavin T
TM7	transmembrane domain 7
TMD	transmembrane domain
TRIAD	transposon-based mutagenesis approach
TTR	human transthyretin
Tyr	tyrosine
UBE2S	Ubiquitin-conjugated E2 S enzyme
UPR	unfolded protein response
Val	valine
vdM	van der Mer
VWF	Willebrand factor
WT	wild-type
XCL1	X-C motif chemokine ligand 1
XFEL	X-ray free electron laser

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

EXECUTIVE SUMMARY

Life essentially depends on proteins and their proper function *in vivo*. To be functionally active and effectively operating, proteins – their polypeptide chains - have to achieve and preserve their three-dimensional structure. The process of how a polypeptide chain achieves a three-dimensional structure is called protein folding. In the cell, this process occurs co-translationally in vectorial fashion right after protein biosynthesis on the ribosomal apparatus. After the synthesis and folding into the 3D structure, the ability of this structure to withstand thermal fluctuations is described by the thermodynamical and/or kinetical parameters, collectively called protein stability. In this report, we applied a systematic search in the selected journals within two years (1st of January 2019 until 14th of January 2021.). The selected journals belong to scientifically most reputable publishing groups – Nature and Science. In the first round of the analysis, we found it helpful to divide publications into two categories: methods and original research papers.

From the analysis of the method papers, the first category, we could clearly distinguish between theoretical and experimental methods. The number of papers describing experimental methods overdubbed the number of papers with theoretical methods. However, it should be noted that while the number of papers referring to theoretical methods is about five times smaller, breakthrough discoveries have been reported by applying AI-based deep learning approaches, as exemplified by the accurate prediction of 3D structures of proteins by the DeepMind group.

The second category, original research papers, can be divided into several sub-categories, such as chaperones, protein degradation, amyloid folding/misfolding, and membrane proteins – as the most dominant topics. Notably, the central drive of our increasingly more extensive knowledge of proteins stems from applying new methods primarily based on protein engineering (deep mutational scanning, for example), single-molecule technique and thermal profiling as the most frequent approaches. In the final section of this report, four selected papers are summarized more in detail.

In summary, this report provides an overview of the current challenges and actual hot topics in protein science. We are convinced that the report can help us focus on specific big questions that the protein community is currently discussing.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

TABLE OF CONTENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

1. METHODOLOGY AND CONCEPTUAL SCHEME FOR THE REPORT PREPARATION

1.1 Selection and analysis of publications

To prepare the report, we decided to use a systematic approach to the literature analysis. In the first round of the selection, we searched for relevant publications, which had appeared in the most prominent journal groups: Science and Nature publishing groups. The reason for selecting these journals/publishers is that we would like to obtain a comprehensive overview of the cutting-edge development in the field of protein folding and stability. Searching keywords with corresponding Boolean operators were: protein AND folding OR stability. These keywords were searched in full articles, reports, protocols, and reviews.

Regarding the timeline, we focused on two years – from 1.12019 until 14.1.2021. Overall, the outcome was 1090 publications; out of these, 209 publications were in Science and related journals, and 881 publications were in Nature publishing groups (for summary, see Fig. 1.1).

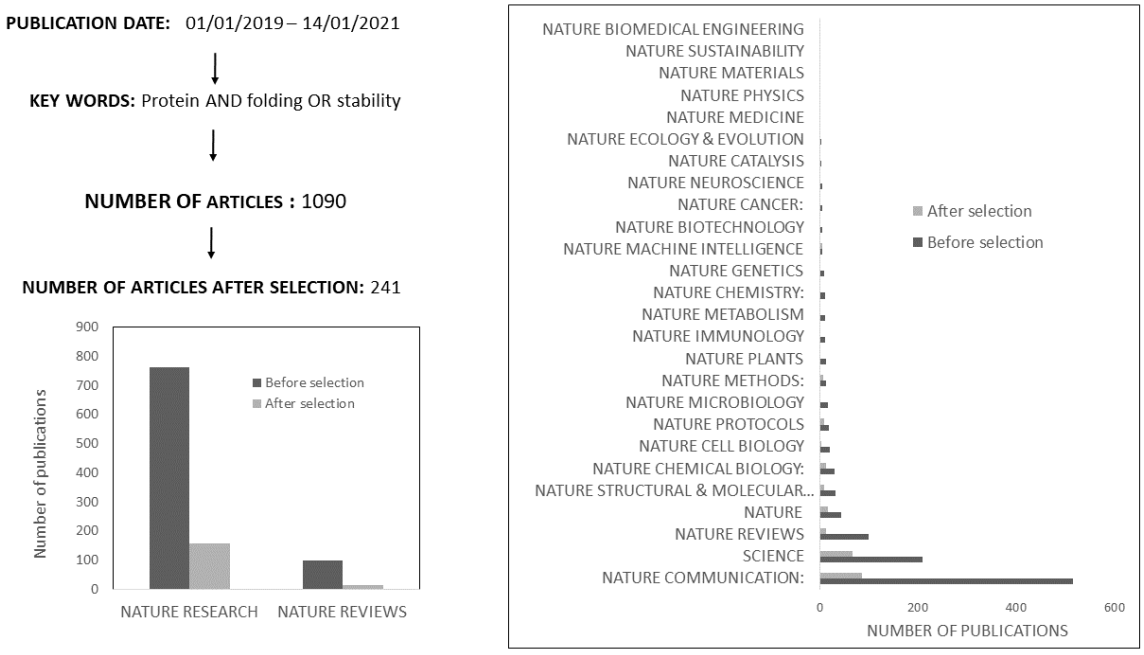

Figure 1.1 Summary of the search strategy for relevant publications in the field of protein science in the Science and Nature publishing group.

Next, we analyzed the Titles and Abstracts of these publications and selected relevant papers. After this selection round, we obtained 241 (22.1%) publications. In general, the publications are distributed

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

in different types of journals. Some of the journals from the Nature group publish only review articles, for this reason we lumped these journals in one group – Nature reviews. This group consists of the following journals, and the number of the appropriate publications is given in parentheses: Nature Protocols (18), Nature Review physics (1), Nature Reviews cancer (8), Nature Reviews Cardiology (3), Nature Reviews Chemistry (7), Nature Reviews Clinical Oncology (2), Nature Reviews Disease Primers (1), Nature Reviews Drug Discovery (10), Nature Reviews Endocrinology (3), Nature Reviews Gastroenterology & Hepatology (4), Nature Reviews Genetics (7), Nature Reviews Immunology (8), Nature Reviews Materials (1), Nature Reviews Methods Primer (1), Nature Reviews Microbiology (9), Nature Reviews Molecular Cell Biology (25), Nature Reviews Nephrology (3), Nature Reviews Neurology (2), Nature Reviews Neuroscience (2), Nature Reviews Physics (1), Nature Reviews Rheumatology (1), Nature Reviews Urology (1). Overall, the most publications after selection were found in Nature communication (87 papers = 36.1 %) followed by Science (66 papers = 27.4 %), Nature (16 papers = 6.6 %) and Nature reviews (14 papers = 5.8 %). We found that manual selection based on Title/Abstract had different consequences in reducing the number of the publications. Even though the publications contained keywords, they were not devoted to the topic and occurred in the publication with a different research focus. Among the first two journals, we found that the manual selection reduced more publication in Nature Communication than in Science. We described this as high/low topic cohesion, which varies between journals. Next, we compared Nature research journals versus Nature reviews and found 14 papers in reviews and 156 papers in research journals after the selection.

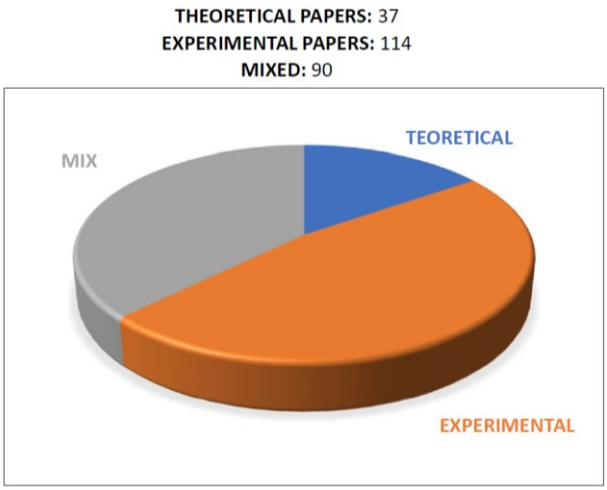

Figure 1.2 After selection, papers were classified into three categories: theoretical, experimental, and mixed publicati

ons.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Next, we found during the manual inspection that publications can be sorted into three categories experimental, theoretical, and mixed. Most of the papers, 114 out of 241, were experimental, 37 were theoretical, and 90 papers contained a mix of experimental and theoretical parts (Fig. 1.2).

This categorization was based on the reading of the abstracts and hence on the major research novelty achieved by experiment or by theory. We did not in-depth analyze whether a theoretical approach was used as a support. We used the word counting tool (https://www.online-utility.org/) to obtain the frequency of the words in the titles and abstracts. From the first top 300 frequent words, we selected those which are informative for the research topic. Non-informative words like“the, which, and,“ were not considered further. Similarly, we omitted the words „protein folding and stability“ because they were searching terms and hence trivial. The word frequency analysis pointed out that the most frequently used word in the title and abstracts of the selected publications is “structure, “ used 220 times, which makes 0.488 % of the total number of words. The word „structure“ is followed by words „interaction“ (140 times = 0.311 %), „design“ (131 = 0.291 %), „complex“ (109 = 0.242 %), „membrane“ (109 = 0.242 %) and „chaperones“ (108 = 0.239 %). The results of the word counting are summarized in Fig. 1.3.

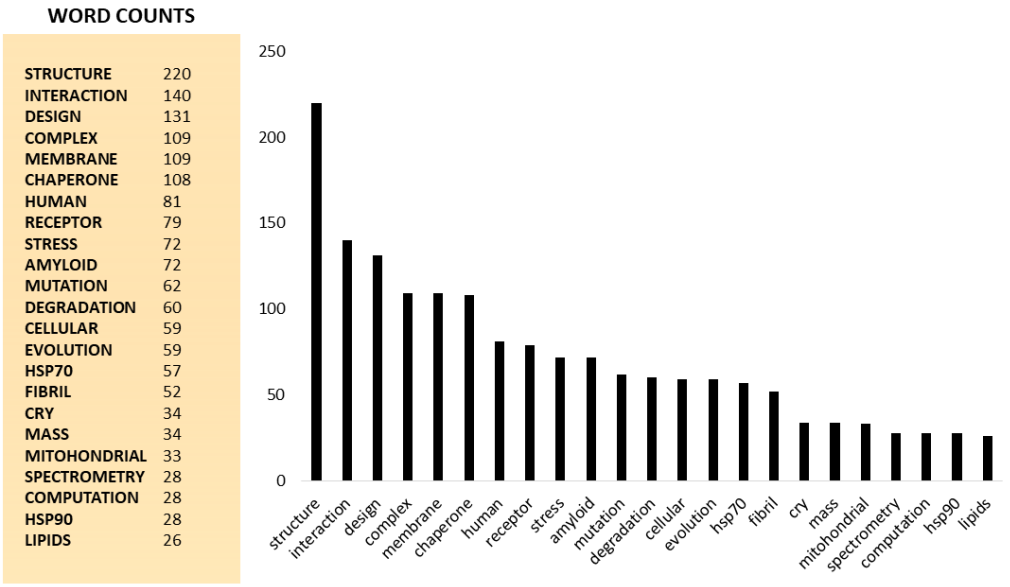

Figure 1.3. Results of the word counting in the titles and abstracts of the selected papers. Further sub-categorization is possible.

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement N° 952333

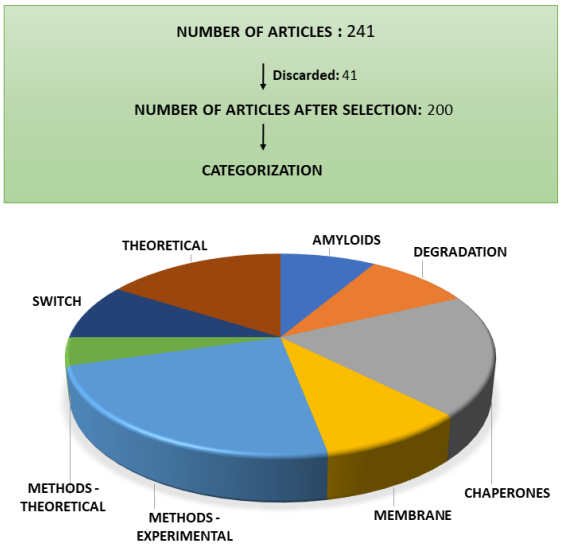

Figure 1.4 The second round of the selection and categorization according to sub-topic and sub-field of the protein folding and stability.

While categorizing publications according to research sub-fields and sub-topics, we found several publications that were difficult to categorize. Hence, we have performed a second round of the selection. In this round, we have discarded 41 papers. Categorization according to the topics showed us eight categories: amyloids, degradation, chaperones, membrane, methods experimental, methods theoretical, switch, theoretical research. In each category, there are a different number of papers.

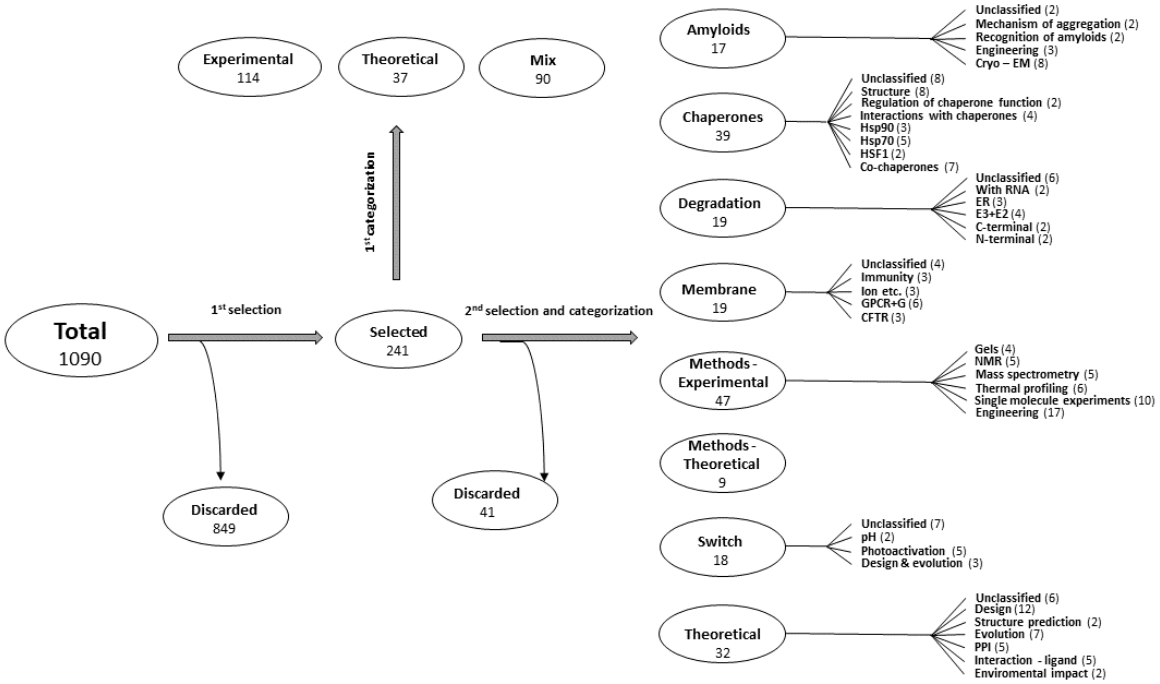

Figure 1.5 Summary of the selection and categorization procedures for publications in protein science.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Experimental methods had the highest number of papers in research journals. Similarly, the majority of the reviews are in the experimental science (48), while only seven papers are considered as theoretical reviews. The chaperone category has 39 papers, followed by 32 theoretical research papers and 19 membrane and degradation papers (see Fig. 1.4). Next, each category can be further sub-categorized according to more specific topics using the combination of word-counting and manual inspections. The final summary of the selection and categorization is presented in Fig. 1.5.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

2. METHODS IN PROTEIN FOLDING AND STABILITY

Our knowledge and understanding of the complex biomacromolecules, their folding, and stability heavily depend on developing new assays and experimental methods. During our analysis of the publications in Science and Nature publisher groups, we found that in many research papers the authors have used or developed novel assays and methods. Therefore, we decided to separate the publication concerning methods and research publications. After separation, we identified six categories of experimental methods and techniques: protein engineering (17 papers), mass spectrometry (MS) (5 papers), single-molecule (10 papers), thermal profiling (6 papers), gels (4 papers), nuclear magnetic resonance (NMR) (5 papers) and new theoretical approaches (9 papers).

2.1 Protein engineering

Most research papers and reviews were devoted to new approaches in the protein engineering fields. Gadella and his group (Bindels *et al.* 2020) describe a screening method for developing optimized red fluorescent proteins. The method consists of two screening rounds. The first round is based on the fluorescence of individual bacterial cells. The second round is based on the screening of mammalian cell lines. Using this protocol, an exceptional bright protein called mScarlet was obtained. In a different approach, Oliinyk *et al.* (Oliinyk *et al.* 2019) developed fluorescent proteins with emission in the near-infrared region. The authors used cyanobacteriochrome that efficiently binds biliverdin chromophore, which led to a version of protein miRFP670nano that can be excited at 645 and 670 nm. Crystal structures of the abovementioned proteins were obtained.

Cafiso (Joseph *et al.* 2019) developed an approach for a high-precision distance analysis of outer-membrane proteins in the nanometer range. The methodology is based on in situ labeling of engineered cysteines (Cys) with a methanethiosulfonate spin-label. Interspin distances are measured in situ, using pulsed electron-electron double resonance (pulsed electron-electron double resonance or double electron-electron resonance) spectroscopy. Seeger (Zimmermann *et al.* 2020) provided a protocol to generate synthetic nanobodies – sybodies against any purified protein or protein complex. The procedure entails a single round of ribosome display, followed by two rounds of phage display and binder identification by enzyme-linked immunosorbent assay. Using fragment exchange cloning

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

methods, the sybody sequences are transferred to expression vectors without amplifying them by polymerase chain reaction.

Understanding of protein folding can be manifested by the ability to design new stable foldable artificial proteins. Along this line, the Tanaka group (Shohei Eda *et al.* 2019) showed the creation of a human serum albumin-based ArM (artificial metalloenzyme) that can protect the catalytic activity of a bound ruthenium metal, even in the presence of 20 mM glutathione. To exploit this biocompatibility, they developed the concept of glycosylated artificial metalloenzymes, which is based on functionalizing ArMs with N-glycans targeting moieties. Brun and his group described a method for labeling and imaging pili and other surface-exposed nanomachines in live cells by using a combination of genetics and traditional maleimide-based click chemistry. A cys substitution is made in the major pilin subunit for subsequent labeling with thiol-reactive maleimide dyes. Brun and his group (Ellison *et al.* 2019) described parameters for selecting cys substitution positions, optimized labeling conditions for epifluorescence imaging of pilus fibers, and methods for impeding pilus activity. Schmieder and Lehner (Schmiedel and Lehner 2019) presented a method that allows to obtain a high-resolution three-dimensional structure of a macromolecule. This method is based on the measurements of the activity of the molecule´s mutant variants. The technique is enabled by quantifying genetic interactions (epistasis) between mutations and discriminating direct from indirect interactions. Hollfelder group (Emond *et al.* 2020) showed a transposon-based mutagenesis approach (TRIAD) to generate libraries of random variants with short in-frame insertions and deletions, and screen TRIAD libraries to evolve a promiscuous arylesterase activity in a phosphotriesterase. Cong *et al.* (Cong *et al.* 2019) investigated coevolution between 5.4 million pairs of proteins in *Escherichia coli* and 3.9 million pairs in *Mycobacterium tuberculosis*. They found strong coevolution complexes involved in metabolism. The combination of this coevolution with structure modeling, to predict protein-protein interactions (PPIs), brings high accuracy. Mann *et al.* (Mann *et al.* 2020) implemented high-throughput-controlled radical polymerization techniques to generate a large library of acrylamide carrier/dopant copolymer (AC/DC) excipients to reduce insulin aggregation. The top-performing AC/DC excipient candidate enabled the development of an ultrafast-absorbing insulin lispro (a fast acting insulin analogue) formulation, which remains stable under stressed aging conditions five times longer than commercial Humalog® (commercial fast acting lispro insulin produced by Eli Lilly and Company, American Pharmaceutical Company). Kuhlman and Bradley (Kuhlman and Bradley 2019) described current approaches for protein structure prediction and design and highlighted a selection of successful applications. They presented new algorithms for designing protein folds and protein-protein interfaces, which were used

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

to engineer novel high-order assemblies and design from scratch fluorescent proteins with novel or enhanced properties.

Domain swapping is the process by which identical monomeric proteins exchange structural elements to generate dimers/oligomers. The application of domain-swapping in protein design has been limited due to the lack of simple and reliable design approaches. Nandwani (Nandwani *et al.* 2019) demonstrated that the hydrophobic "cystatin motif" (QVVAG) from the domain-swapping protein Stefin B, when engineered into a solvent-exposed, tight surface loop between two β-strands prevents the loop from folding back upon itself and drives domain swapping in any protein, i.e. originally non-domain-swapping proteins. Sang *et al.* (Sang *et al.* 2020) reported the exploration of helical sulfone-γ-AApeptides with entire unnatural backbones for their ability to structurally and functionally mimic glucagon-like peptide1 (GLP-1). The results suggest that efficient construction of novel GLP-1 receptor agonists could be achieved with nanomolar potencies and sulfono-γ-AApeptides display a higher stability against enzymatic degradation compared to GLP-1. Yan group (Yu, Zhao, *et al.* 2020) explored the possibility of developing mechanically stable heterodimerization systems from a set of recently engineered helix-heterotetramers. They randomly chose two helix-heterotetramers and modulated the mechanical properties by changing the stretching geometry and the number of interacting helices. Next, theyappliedd helix-heterotetramers in single-molecule manipulation studies temperature-dependentpendent unfolding and refolding of a titin immunoglobulin domain and α-actinin spectrin repeats. Mikoshiba group (Kabayama *et al.* 2020) reported a versatile method for engineering an ultra-stable cytoplasmic antibody, with a strong estimated net negative charge at pH 6.6, by fusing peptide tags. The key feature of this tag is that they are enriched by acidic amino acids and hence have a highly negative charge and a low isoelectric point.

Benhaim *et al.* (Benhaim *et al.* 2020) demonstrated how concurrent reorganizations at the hemagglutinin (HA) 1 receptor-binding domain (RBD) interface and HA2 fusion subunit produce a dynamic fusion ensemble in full-length HA. By using pulse-labeling hydrogen/deuterium exchange MS and cryoelectron tomography to monitor and characterize the structural dynamics of HA during fusion activation on intact virions, they resolved that the soluble HA ectodomain is transformed directly to the postfusion state with no observable intermediate. Rollins *et al.* (Rollins *et al.* 2019) described an experimental method of three-dimensional (3D) structure determination that exploits the increasing ease of high-throughput mutational scans. The authors analyzed five large-scale mutational scans and discovered that the pairs of residues with the largest positive epistasis in the experiments are sufficient to determine the 3D fold. From genetic screens of three proteins, the authors showed that the strongest epistatic pairings and revealed 3D contacts between molecules.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

2.2 Development in single-molecule methods

Kaiser group (Chen, Rajasekaran, *et al.* 2020) combined *in vivo* and single-molecule experiments and followed the co-translational folding of the G-domain. The results suggest a strictly sequential folding pathway initiating from the C-terminus. Folding and synthesis proceed in opposite directions. Schwalbe *et al.* (Schulte *et al.* 2020) studied the creation of disulfide bonds in eye-lens protein γB-crystallin in the ribosomal exit tunnel. Fernandez and his group (Alonso-Caballero *et al.* 2021) studied protein Cpa, which is utilized by *Streptococcus pyogenes* for adhesion on surfaces. Using magnetic tweezers force spectroscopy, they demonstrated that when Cpa folded at low forces (<6pN), the Cpa thioester bond reacts reversibly with amine ligands, common in inflammation sites. Mechanical unfolding and exposure to forces >6pN block thioester reformation. Choi *et al.* (Choi *et al.* 2019) describe a single-molecule force microscopy technique for monitoring the folding of helical membrane proteins in vesicle and bicelle environments. They characterized *Escherichia coli* rhomboid protease GlpG and human β2-adrenergic receptor. Both proteins fold in a strict N-to-C-terminal fashion and create units of helical hairpins. Tapia-Rojo *et al.* (Tapia-Rojo, Alonso-Caballero, and Fernandez 2020) measured the interaction between the vinculin head and the talin R3 domain. Binding events start as a short contraction of the unfolded talin polypeptide, what is caused by the reformation of the vinculin-binding site helices, this dictates that this reaction is regualated by biphasic mechanism. Talin-vinculin-actin association could operate as a negative feedback mechanism to stabilize force on focal adhesion. Liu *et al.* (Liu, Liu, *et al.* 2020) reported the molecular mechanism of an ultrastable protein complex responsible for resisting shear forces and adhering bacteria to cellulose fibers in the human gut. They used the mechanically ultrastable *R. champanellensis* dockerin: cohesin complex to explain mechanistically how gut microbes regulate cell adhesion strength at high shear stress. Serdiuk *et al.* (Serdiuk *et al.* 2019) studied insertion and folding pathways of single membrane proteins guided by translocases and insertases. They used YidC and SecYEG for initiation folding of the unfolded polypeptide by inserting a single structural segment. YidC inserts the segments in random order, whereas SecYEG inserts them sequentially. When YidC and SecYEG cooperate, the folding pathway is dominated by the translocase. They proposed that both fundamentally different pathways along which YidC and SecYEG insert and fold a polypeptide are essential components of the membrane protein biogenesis.

2.3 Thermal profiling methods

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Moellering group (Huang *et al.* 2019) described a proteomic method - Hotspot Thermal Profiling- to detect the effects of site-specific protein phosphorylation on the thermal stability of native proteins in live cells. This method can detect intrinsic changes to protein structure and extrinsic changes to protein-protein and protein-metabolite interactions. Bergamini group (Perrin *et al.* 2020) introduced tissue thermal proteome profiling for measuring the binding of small-molecule drugs to proteins in tissue samples from drug-treated animals by detecting changes in protein thermal stability by quantitative MS. This method will help to elucidate the mechanism of drug action *in vivo*. Bantscheff group (Kalxdorf *et al.* 2021) introduced cell surface thermal proteome profiling to characterize ligand-induced changes in proteins and thermal stabilities at the plasma membrane. They demonstrated drug binding to extracellular receptors and transporters, discovered stimulation-dependent remodeling of T-cell receptor complexes, and described a competition-based approach to measure target engagement o G-protein-coupled receptor (GPCR) antagonists. Jarzab *et al.* (Jarzab *et al.* 2020) used a MS-based proteomic approach to compile an atlas of the thermal stability of 48 000 proteins from 13 species ranging from *Archae*a to humans and covering melting temperatures of 30-90°C. They found that protein sequence, composition, and size affect in prokaryotes and eukaryotic proteins. Moreover, they showed a nonlinear relationship between the degree of disordered protein structure and stability. Savitski group (Mateus *et al.* 2020) used thermal proteome profiling to measure the proteome-wide protein abundance and thermal stability in response to 121 genetic perturbations in *E. coli*. They found that functionally associated proteins have coordinated changes in abundance and stability across perturbations, owing to their co-regulation and physical interactions.

Bozdech and his group (Dziekan *et al.* 2020) demonstrated the cellular thermal shift assay (CETSA) protocol in two variants - the melt curve and the isothermal dose-response, representing a strategy for identifying antimalarial drug targets. CETSA enables proteome-wide target screening for unmodified antimalarial compounds with an undetermined mechanism of action, providing quantitative evidence about direct drug-protein interactions.

2.4 Mass spectrometry methods

The Fornasiero group (Alevra *et al.* 2019) described MS protocol for quantitative proteome turnover analysis in mice based on a commercially available diet for stable isotope labeling of amino acids in mammals. They cover the practical aspect of metabolic labeling and explain both the experimental and computational steps that must be taken to obtain accurate *in vivo* results. They present a simple experimental workflow that enables measurement in a time frame of 4-5 weeks,

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

including the labeling time. Sharon group (Vimer, Ben-Nissan, and Sharon 2020) described a native MS approach to rapidly characterize intact overexpressed proteins immediately from crude samples. They demonstrate the methods applicability for determining activity and interaction of regulatory protein and for glycosylation analysis of a designed antibody. The authors described a rapid and straightforward protocol, which can be carried out on multiple mass spectrometers, relevant to both prokaryotic and eukaryotic expression systems,. Zenobi group (Marchand *et al.* 2020) reported the design of a temperature-jump electrospray source for MS that allows to perform fast kinetics experiments (0.16-32 s) at different temperatures (10-90 °C). The setup allows recording folding and unfolding kinetics by using temperature jumps from high-to-low and low-to-high temperatures. Politis group (Martens *et al.* 2019) described a protocol for uncovering key lipid-protein interactions implicated in stabilizing important functional conformations of membrane proteins. This method uses hydrogen-deuterium (H/D) exchange mass spectrometry (HDX-MS) of membrane proteins for obtaining information on the lipid species. Then, by molecular dynamics (MD) simulations in the lipid bilayer, lipid-protein interactions have been revealed and experimentally tested by HDX-MS. Gross labs (Liu, Rempel, and Gross 2020) described the fast photochemical oxidation of proteins, the conditions for successful footprinting and its examination by mass measurements, post-labeling sample handling and digestion, the liquid chromatography-tandem MS analysis of the digested sample and the data analysis with Protein Metrics Suite.

2.5 Development in protein NMR methods

Williamson group (Fowler, Sljoka, and Williamson 2020) presented a method that measures the accuracy of NMR protein structures. It compares random coil index against local rigidity predicted by mathematical rigidity theory, calculated from NMR structures, using a correlation score (which assesses secondary structure) and a root-mean-square deviation score (which measures overall rigidity). Comparing NMR to crystal structures shows that secondary structure is equally accurate, but crystal structures are typically too rigid in loops, whereas NMR structures are typically too floppy overall.

Xiao *et al.* (Xiao *et al.* 2019) described the methodology for monitoring membrane protein unfolding in lipid membranes. This method is based on H/D exchange and SSNMR to detect changes in the protein water-accessible surface and monitor the reversibility of unfolding. They applied the H/D exchange methodology on a seven-helical photoreceptor for obtaining the atomistic description. Orlando *et al.* (Orlando, Raimondi, and W 2019) propose an autoencoder-based method – ShiftCrypt,

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

which offers a concept for analyzing, comparing, and interpreting CS (chemical shifts) of proteins in their native, multidimensional space. CS brings information about the conformation of the protein and the movements it performs. CS identify hidden similarities between diverse proteins and peptides and differences between the same protein in two different binding sites. Kovermann (Kohn and Kovermann 2020) described the multidimensional high-resolution NMR spectroscopy method. It used a solid base, which can be expanded to binding partners to recognize functional consequences of crowded conditions. By using this method, the thermodynamic stability of a beta-barrel protein was studied. Forman-Kay group (Dawson *et al.* 2020) shows how non-cooperative folding and conformational exchange leads to a graded inhibition of 4E-BP2:eIF4E binding, shifting 4E-BP2 into an eIF4E binding incompatible conformation and regulating translation initiation. C-terminal intrinsically disordered region (C-IDR) is non-cooperatively folded, with an exchange between beta-strands and helical conformation. Phosphorylation of C-IDR shifts the conformation equilibrium, controlling access to eIF4E binding sites.

The authors used MS, theoretical simulations, nuclear polarization-enhanced solid-state NMR (SSNMR) and cryo-electron microscopy to show that cys residues undergo S-glutathionylation and S-nitrosylation and form non-native disulfide bonds, which can be used for guiding protein folding. Stone *et al.* (Stone *et al.* 2019) studied the capsid structure of thermostable phage P74-26 with cryoelectron microscopy (cryo-EM). The architecture of the capsid has T=7 geometry despite being twice as large as mesophilic homologs. Based on these results, they predict that decreased icosahedral complexity (T ≤7) leads to a more stable capsid assembly. Butera and Hogg (Butera and Hogg 2020) tested the redox state of disulfide bonds of fibrinogen from human donors. They identified that fibrinogen is produced in multiple disulfide-bonded or covalent states. These states are dynamic because they are changed by fluid shear forces *ex vivo* and *in vivo*. They demonstrate that fibrinogen exists and functions in multiple covalent forms. Feng *et al.* (Feng, Gruebele, and Davis 2019) demonstrated a pipeline that integrates meganuclease-mediated mosaic transformation with fluorescence-detected temperature-jump microscopy to probe the dynamics and stability of endogenously expressed proteins in different tissues of living multicellular organisms.

2.6 Protein gel-based methods

The application of the principles of protein folding have emerged at macroscopic scales– the production of protein-gels. Khoury and Popa (Khoury and Popa 2019) used programmed protein hydrogels for developing protein-based biomaterials. They used protein-polymer interactions to

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

manipulate the stiffness of protein-based hydrogels made from bovine serum albumin by using polyelectrolytes - polyethyleneimine and poly-L-lysine. Tsurusawa *et al.* (Tsurusawa, Arai, and Tanaka 2020) found a new type of gelation in dilute colloidal suspensions, in which percolation occurs after the local formation of mechanically stable, rigid clusters. Topological percolation generates little mechanical stress, and the resulting gel is almost stress-free when formed. Chen *et al.* (Chen *et al.* 2019) reported a method for preparing single-chain nanogels (SCNGs) with improved efficiency. They also investigated the impact of the dynamic molecular conformational change of SCNGs on cellular interactions. These single-chain nano-objects can be used in the biomedical area. Khoury *et al.* (Khoury *et al.* 2020) reported a new method to program protein hydrogels and induce shape changes in aqueous solutions. They used hydrogels from serum albumin, which are synthesized in a cylindrical or flower shape. The gels are then programmed into a spring or a ring shape, and programming is performed through a marked change in stiffness induced by the adsorption of Zn^{2+} or Cu^{2+} cations.

2.7 Theoretical methods used in protein folding and stability

One of the most hottest development in theoretical methods include application of deep learning and several algorithms inspired from natural language processing field. A highly accurate protein structure prediction was achieved using artifial intelligence program so-called AlphaFold (Jumper *et al.* 2021) . At the heart of the latest version of AlphaFold is a new machine learning approach that incorporates physical and biological knowledge of protein structure into the design of a deep learning algorithm using multiple sequence alignments. The neural network AlphaFold that DeepMind team developed was entered into the so-called CASP14 assessment. Critical Assessment of protein Structure Prediction (CASP) is a community-wide, experiment for protein structure prediction that takes every two years. CASP provides research groups with a chance to objectively validate their structure prediction methods. The Alphafold authors developed the first computational approach capable of predicting protein structures in some cases to near experimental accuracy.

Rational protein engineering requires a full understanding of protein function. The Church group (Alley *et al.* 2019) described the application of deep learning to unlabeled amino acid sequences to extract information about the basic properties of a protein. The authors achieved a statistical representation of a protein that is sematically rich and based on structure, evolution, and biophysical properties. Based on this unified representation (UniRep), researcher will able to predict the stability of natural and de novo designed proteins. UniRep further enables two orders of magnitude efficiency improvement in a protein engineering task.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

A new pipeline for rapid protein structure prediction was developed (Wenzhi Mao *et al.* 2020). It consists of two tools: a predictor of a contact between residues named as AmoebaContact, and a contact-assisted folder, GDFold. The combination and cooperation of these two tools increase the predictive power. A general strategy was developed based on the principle of digital reading. The strategy allows efficiently obtaining enzyme mutants with the desired activities (Zhang *et al.* 2019).

De novo design of inhibitors against coronaviruses was employed using two different approaches (Cao *et al.* 2020). Computer-generated scaffolds were either targeting the ACE2 helix of human receptor that interacts with the spike RBD or scaffolds were directly targeting RBD. Their amino acid sequences have been designed to optimize target binding, folding, and stability. These hyper-stable mini binders belong to novel therapeutics for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Pearce *et al.* (Pearce *et al.* 2019) developed a flexible and robust numerical framework for deriving Markov transition networks directly from time-independent data that have been sampled from stationary equilibrium distributions. They demonstrated the potential of their approach by reconstructing network dynamics for protein folding, various motifs of gene-regulatory networks, and developmental pathways of HIV. Network topology prediction agrees well with a direct estimates from the time-dependent MD data, stochastic simulations, and phylogenetic trees.

Systematic research of transient path-time symmetry and evidence of its decay from the equilibrium state in the molecular and intermediate scale was described (Gladrow *et al.* 2019). In automated experiments with Brownian dynamics, they determine the symmetry of colloidal pulses driven by fNewton forces. The results capture a single coordinate in a multidimensional free energy environment, such as those found in electrophysiology and single-molecule experiments using fluorescence. A workflow for small-scale pre-crystallization screening was described (Audet *et al.* 2020). It involves detecting GPCR expression levels in *Spodoptera frugiperda* (Sf9) culture by flow cytometry and the evaluation of GPCR stability. Penn *et al.* (Penn *et al.* 2020) have recently found that rhodopsin expression and maturation are limited by the hydrophobicity of its seventh transmembrane domain (TM7), which contains polar residues that are essential for function. Based on these observations, they hypothesized that rhodopsin expression should be less tolerant of mutations in TM7 compared to mutations in the hydrophobic transmembrane domain (TMD). In conjunction with evolutionary analysis, these results suggest that solvation energy is likely to limit the evolutionary space of the sequence of polar TM domains. In an in-depth study of 100 taxonomically different organisms , a robust proteomic workflow was developed (Muller *et al.* 2020). In this workflow, the step of peptide separation is performed by a microstructured and extremely reproducible chromatographic system. The data also provide an extensive case study for a machine learning, as demonstrated by

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

experimental validation of the predicted properties of peptides from *Bacteroides uniformis*. The results offer a comparative view of the functional organization of organisms throughout the evolutionary range. As a rule, proteins and proteomes are remarkably diverse between organisms. One can easily explore and functionally compare using the web-service at www.proteomesoflife.org.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

3. THEORETICAL RESEARCH ON PROTEIN STABILITY AND FOLDING

3.1 Evolution of proteins

Pillai *et al.* (Pillai *et al.* 2020) showed that hemoglobin evolved from an ancient monomer and characterized the historical 'missing link' through which the modern tetramer evolved. First, a noncooperative homodimer with high oxygen affinity existed before the gene duplication that later generated distinct α- and β-subunits. This findings confirmed that evolution can produce new complex molecular structures and functions via simple genetic mechanisms that recruit existing biophysical features into higher-level architectures.

The work of Spohn *et al.* (Spohn *et al.* 2019) indicates that the evolution of resistance against certain antimicrobial peptides (AMPs), such as tachyplesin II and cecropin P1, is limited. The resistance level provided by point mutations and gene amplification is very low, and antibiotic-resistant bacteria display no cross-resistance to these AMPs. They found that simple physicochemical features can define the propensity of bacteria to evolve resistance against AMPs. This work could serve as a promising source for developing new AMP-based therapeutics eith a less prone resistance development.

Hypotheses about the origin of eukaryotic cells are classically framed within the context of a universal 'tree of life' based on conserved core genes. Williams *et al.* (Williams *et al.* 2020) conducted bioinformatic analysis to reevaluate the evidence for two-domain tree and applied large amount of data by using supertree and coalescent methods to interrogate >3,000 gene families in *Archae*a and eukaryotes. They found that eukaryotes consistently originate from within the *Archae*a in a two-domain tree. The authors' analyses support a close relationship between eukaryotes and *Archae*a and identify the *HeimdallArchaeota* as the current best candidate for the closest *Archae* relatives of the eukaryotic nuclear lineage.

Protein sequences contain rich information about protein evolution, fitness landscapes, and stability. The Brooks group (Ding, Zou, and Brooks Iii 2019) investigated how latent space models trained using variational auto-encoders can infer these properties from sequences. Using both simulated and real sequences - low dimensional latent space representation of sequences, calculated using the encoder model, captures evolutionary and ancestral relationships between sequences. With experimental fitness data and Gaussian process regression, the latent space representation also enables learning the protein fitness landscape in a continuous low dimensional space. The model is also useful in predicting protein mutational stability landscapes and quantifying the importance of

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

stability in shaping protein evolution. Hence, these models are well-suited to guide protein engineering efforts.

Natural selection can promote or hinder a population's evolvability-the ability to evolve new and adaptive phenotypes-but the underlying mechanisms are poorly understood. Wagner group (Zheng, Guo, and Wagner 2020) subjected populations of yellow fluorescent protein to directed evolution under different selection regimes and then evolved them toward the new phenotype of green fluorescence. The experiments show how selection can enhance evolvability by enhancing robustness and create the conditions necessary for evolutionary success.

Adaptation, where a population evolves increasing fitness in a fixed environment, is typically thought of as a hill-climbing process on a fitness landscape. In general, it is not clear how stochastic tunneling across fitness valleys would affect long-term fitness evolution. Using a spin-glass type model for the fitness function that considers microscopic context-dependence of the effects of mutations (Guo, Vucelja, and Amir 2019), authors found that hopping between metastable states can mechanistically and robustly give rise to a slow fitness trajectory.

Understanding the pattern of the non-independence of mutations is critical for relating genotype and phenotype. However, the combinatorial complexity of potential epistatic interactions has severely limited the analysis of this problem. Ranganathan group (Poelwijk, Socolich, and Ranganathan 2019) reported a comprehensive experimental study of 213 mutants that link two phenotypically distinct variants of the quadricolor fluorescent protein. The data show the existence of many high-order epistatic interactions between mutations and reveal extraordinary sparsity, enabling novel experimental and computational strategies for learning the relevant epistasis. They demonstrate that such information can be used to predict phenotypes accurately even with limited experimental data.

3.2 Protein design and theory

Peacock (Peacock 2020) shared a perspective on a protein design. How and why proteins fold from peptide sequences to yield a particular structure, have continued for decades and have inspired efforts to design proteins *de novo*—that is, to design structured miniature protein folds from first principles rationality. Koepnick *et al.* (Koepnick *et al.* 2019) developed *de novo* protein design for the online protein-folding game called Foldit. Players were presented with a fully extended peptide chain and challenged to craft a folded protein structure and an amino acid sequence encoding that

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

structure. Foldit players can now—starting from an extended polypeptide chain—generate a diversity of protein structures and sequences that encode them *in silico*. 146 Foldit player designs with sequences unrelated to naturally occurring proteins were encoded in synthetic genes; 56 were found to be expressed and soluble in *Escherichia coli* and to adopt stable monomeric folded structures in solution. Foldit makes possible to apply the hidden knowledge that contributes to success in *de novo* protein design and shows that non-professional scientists can discover creative new solutions to outstanding scientific challenges such as the protein design problem. Correia group (Sesterhenn *et al.* 2020) developed a protein design algorithm called TopoBuilder. With TopoBuilder, they engineered epitope-focused immunogens displaying complex structural motifs. In both mice and nonhuman primates, cocktails of three *de novo*–designed immunogens induced robust neutralizing responses against the respiratory syncytial virus. The immunogens refocused preexisting antibody responses toward defined neutralization epitopes. The design approach opens the possibility of targeting specific epitopes for the development of vaccines and therapeutic antibodies. Jalan *et al.* (Jalan *et al.* 2020) designed three triple-helical heterotrimers that each contains a putative von Willebrand factor (VWF), and discoidin domain receptor (DDR) recognition sequence from COL1, which were designed with chain B permutated in all three positions. AAB demonstrated a strong preference for both VWF and DDR and also induced higher levels of cellular DDR phosphorylation. It showed that COL1 adopts an AAB register. Pan *et al.* (Pan *et al.* 2020) developed a computational design method, loop-helix-loop unit combinatorial sampling (LUCS), that mimics nature's ability to create families of proteins with the same overall fold but precisely tunable geometries. Biophysical characterization showed that 17 (38%) of 45 tested LUCS designs encompassing two different structural topologies were well folded, including 16 with designed non-native geometries. Four experimentally solved structures closely matched the designs. LUCS greatly expands the designable structure space and offers a new paradigm for designing proteins with tunable geometries that may be customizable for novel functions. DeGrado group (Polizzi and DeGrado 2020) developed a unit of protein structure—a van der Mer (vdM)—that maps the backbone of each amino acid to statistically preferred positions of interacting chemical groups. They designed six *de novo* proteins using vdM to bind the drug apixaban. The vdMs may design proteins for sensing, medicine, and catalysis. Baker labs (Chen, Kibler, *et al.* 2020) described the design of two-input AND, OR, NAND, NOR, XNOR, and NOT gates built from *de novo*–designed proteins. These gates regulate the association of arbitrary protein units ranging from split enzymes to transcriptional machinery *in vitro*, in yeast, and in primary human T cells, where they control the expression of the TIM3 gene related to T cell exhaustion. The modularity and cooperativity of the control elements, coupled with the ability to *de novo* design an essentially unlimited number of protein components, should enable the design of sophisticated posttranslational control logic over a wide range of biological

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

functions. Correia group (Yang *et al.* 2021) developed a bottom-up approach to build *de novo* proteins tailored to accommodate structurally complex functional motifs. They design five folds for four distinct binding motifs, including a biofunctionalized protein with two motifs and *de novo* proteins that were functional as components of biosensors to monitor antibody responses and as orthogonal ligands to modulate synthetic signaling receptors in engineered mammalian cells. Their work demonstrates the potential of bottom-up approaches to accommodate complex structural motifs, which will be essential to endow *de novo* proteins with elaborate biochemical functions, such as molecular recognition or catalysis. Boyken *et al.* (Boyken *et al.* 2019) presented a general strategy to design pH-responsive protein conformational changes by reorganizing histidine residues in buried hydrogen-bond networks. They designed homotrimers and heterodimers stable above pH 6.5 but undergo cooperative, large-scale conformational changes when the pH is lowered, and electrostatic and steric repulsion build up as the network histidine residues become protonated. The results demonstrate that environmentally triggered conformational changes can now be programmed by *de novo* protein design. Sun & Marelli (Sun and Marelli 2020) presented templated crystallization of structural proteins to nanofabricate hierarchically structured materials up to a centimeter-scale, using silk fibroin as an example. Silk polymorphs can be engineered by varying the peptide seeds used. Modulation of the concentration between silk fibroin and peptide seeds, silk fibroin molecular weight, and pH allows control over nanofibrils morphologies and mechanical properties. The approach faciles integration of the bottom-up templated crystallization with emerging top-down techniques that enables the generation of macroscopic nanostructured materials with potential applications in information storage/encryption, surface functionalization, and printable three-dimensional constructs of customized architecture and controlled anisotropy.

3.3 Protein-protein and protein-ligand interactions

Wang *et al.* (Wang, Cang, and Wei 2020) used element and site-specific persistent homology (a new branch of algebraic topology) to simplify the structural complexity of protein-protein complexe. They also proposed a new deep learning algorithm called NetTree to take advantage of convolutional neural networks and gradient-boosting trees representation and NetTree for predicting PPI $\Delta\Delta G$ (the change of the free energy change of protein-protein interactions). Tests on benchmark datasets indicate that the proposed topology-based network tree is an essential improvement over the current state of the art in predicting $\Delta\Delta G$. Glasgow *et al.* (Glasgow *et al.* 2019) described a generalizable computational strategy for designing sensor-actuator proteins by building binding sites *de novo* into heterodimeric protein-protein interfaces and coupling ligand sensing to modular actuation through

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

split reporters. They developed protein sensors that respond to farnesyl pyrophosphate, a metabolic intermediate in producing valuable compounds. This computational design strategy opens broad avenues to link biological outputs to new signals. Dokholyan group (Wang *et al.* 2020) developed a computationally efficient network-based method, Ohm, to identify and characterize allosteric communication networks within proteins. Ohm relies solely on the structure of the protein of interest. The use of Ohm to map allosteric networks in a dataset composed of 20 proteins experimentally identified to be allosterically regulated. Webserver Ohm.dokhlab.org automatically determines allosteric network architecture and identifies critical coupled residues within this network. Froning *et al.* (Froning *et al.* 2020) used molecular modeling to identify mutations in the T-cell receptor (TCR) constant domains (Cα/Cβ) that have: (*i*) increased the unfolding temperature of Cα/Cβ by 20 °C, (*ii*) improved the expression of four separate α/β TCRs by 3- to 10-fold, and (*iii*) improved the assembly and stability of TCRs with otherwise poor intrinsic stability. The improved stability and folding of the TCRs reduce glycosylation, perhaps through conformational stabilization that restricts access to N-linked glycosylation enzymes. Gainza *et al.* (Gainza *et al.* 2020) worked on molecular surface interaction fingerprinting (MaSIF), a conceptual framework based on a geometric deep learning method to capture important fingerprints for specific biomolecular interactions. The showcase MaSIF with three prediction challenges: protein pocket-ligand prediction, protein-protein interactions prediction, and ultrafast scanning of protein surfaces to predict protein-protein complexes.

Chen *et al.* (Chen, Chen, *et al.* 2020) developed a framework to quantify biomolecular energy landscape at atomic resolution by examining the statistics of the energy changes that occur when the local environment of a site is changed. The location of patches of highly frustrated interactions correlates with key biological locations needed for physiological function. Atomistic frustration analysis provides a route for screening of more specific compounds for drug discovery. Stohr and Tkatchenko (Stohr and Tkatchenko 2019) used an explicit quantum-mechanical approach of density-functional tight-binding combined with the many-body dispersion formalism and demonstrate the relevance of many-body van der Waals forces both to protein energetics and to protein-water interactions. In contrast to commonly used pairwise approaches, many-body effects substantially decrease the relative stability of native states in the absence of water. Upon solvation, the protein-water dispersion interaction counteracts this effect and stabilizes native conformations and transition states.

Sedlak *et al.* (Sedlak *et al.* 2020) found that the forces required to break the SA/biotin bond depend strongly on the attachment geometry. With atomic force microscope (AFM)-based single-molecule force spectroscopy (SMFS), they measured unbinding forces of biotin from different SA subunits to range from 100 to more than 400 pN and carried out hundreds of all-atom steered MD

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

simulations for the entire system, including molecular linkers. The aforementioned approach revealed the molecular mechanism that causes a fourfold difference in mechanical stability: Certain force-loading geometries induce conformational changes in SA's binding pocket lowering the energy barrier, which biotin has to overcome to escape the pocket. Ou *et al.* (Ou *et al.* 2020) used MD simulations to probe the conformations of adhesive protein Pvfp-5β and its salt-tolerant underwater adhesion on superhydrophilic mica. These pairs can lead to firm surface binding. Simulations suggest that the unmodified tyrosine (Tyr) shows similar functions on surface adhesion by forming a pairing structure with a positively charged residue. The simulations confirmed the presence of these residue pairs and verify the strong binding ability of unmodified proteins using NMR spectroscopy and lap shear tests.

To identify functionally relevant parts of proteins, Zu Belzen (Upmeier zu Belzen *et al.* 2019) showed that occlusion-based sensitivity analysis could leverage the knowledge in deep-neural-network-based protein sequence classifiers. It is the tirst validated approach by successfully predicting positions that mediate small-molecule binding or catalytic activity across different protein classes. Next, they inferred the impact of point mutations on the activity of ERK and HRas. Finally, this approach identified engineering hotspots in CRISPR–Cas9 and anti-CRISPR protein AcrIIA4. The work demonstrates how implicit knowledge in neural networks can be harnessed for protein functional dissection and protein engineering.

3.4 Other publications

The Wu group (Noe *et al.* 2019) developed Boltzmann generators by combining deep learning and statistical mechanics, which are shown to generate unbiased one-shot equilibrium samples of representative condensed-matter systems and proteins. Boltzmann generators use neural networks to learn a coordinate transformation of the complex configurational equilibrium distribution to a distribution that can be easily sampled. Accurate computation of free-energy differences and discovery of new configurations are demonstrated, where thestatistical mechanical tools can avoid rare events during sampling without prior knowledge of reaction coordinates.

Louros *et al.* (Louros *et al.* 2020) took high-resolution structural information on amyloid cores to implement a machine learning approach named Cordax (https://cordax.switchlab.org) that explores amyloid sequence beyond its current boundaries. Clustering by t-distributed stochastic neighbor embedding shows how the approach expanded away from hydrophobic amyloid sequences towards clusters of lower aliphatic content and higher charge, or regions of helical and disordered

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

propensities. These clusters uncouple amyloid propensity from solubility representing sequence flavors compatible with surface-exposed patches in globular proteins, functional amylloids or sequences associated with liquid-liquid phase transitions.

Wang *et al.* (Tong Wang *et al.* 2019) showed how a high-quality fragment library could be built using deep contextual learning techniques. The algorithm, called DeepFragLib bidirectional extended short-term-memory recurrent neural networks with knowledge distillation for initial fragment classification, followed by an aggregated residual transformation network with cyclically dilated convolution for detecting near-native fragments. DeepFragLib improves the position-averaged proportion of near-native fragments by 12.2% over existing methods.

Sočan *et al.* (Socan, Purg, and Aqvist 2020) studied cold-adapted enzymes from psychrophilic species show the general characteristics of being more heat-labile and having a different balance between enthalpic and entropic contributions to the free energy barrier of the catalyzed reaction compared to mesophilic orthologs. Among cold-adapted enzymes, some examples show an enigmatic inactivation at higher temperatures before the unfolding of the protein occurs. They analyze these phenomena by extensive computer simulations of the catalytic reactions of psychrophilic and mesophilic α-amylases. This result allows us to examine the structural basis of thermal inactivation, which turns out to be caused by the breaking of a specific enzyme-substrate interaction. Li *et al.* (Li *et al.* 2021) develop and apply a Bayesian modeling approach to resolve the temperature effects in genome-scale metabolic models (GEM). The approach minimizes uncertainties in enzymatic thermal parameters and dramatically improves the predictive strength of the GEMs.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

4. EXPERIMENTAL RESEARCH ON PROTEIN STABILITY AND FOLDING

4.1 Chaperones and their role in protein folding and stability

Collier *et al.* (Collier *et al.* 2019) studied interaction and up-regulation between actin-binding protein filamin C (FLNC) and molecular chaperone heat shock protein family B Member 1 (HspB1) in mouse models of biomechanical stress. They showed that phosphorylation of HspB1 leads to increased exposure of the residues surrounding the HspB1 phosphosites and facilitates their binding to the multidomain region of FLNC. They found that the extension trajectory of FLNC is modulated by the phosphorylated region of HspB1. Wu *et al.* (Wu, Stull, *et al.* 2019) studied the folding of protein SH3 into the native state in the presence of adenosine triphosphate (ATP)-independent chaperone Spy. They found that the more tightly Spy binds to protein, the more it slows the folding rate of the bound protein. This finding is opposite to their previous research of folding of protein Im7 in the presence of Spy. The efficient chaperone-mediated folding appears to represent an evolutionary balance between interactions of sufficient strength to mediate folding and interactions that are too tight, which tend to inhibit folding. Bock group (Wu, Meyer, *et al.* 2019) showed that GUN1 (a protein of unknown function genome uncoupled) regulates chloroplast protein import through interaction with the import-related chaperone cpHSC70-1. They found that overaccumulation of unimported preproteins in the cytosol causes a GUN phenotype in the wild-type (WT) background. They identified the cytosolic heat shock protein 90 (HSP90) chaperone complex, which is a central regulator of photosynthetic gene expression. Rizzolo *et al.* (Rizzolo *et al.* 2021) studied the cooperation of the ribosome-associated trigger factor (TF) chaperone with the ClpXP degradation complex. They found that the effect of TF on ClpXP-dependent degradation varies based on the nature of the substrate. TF increases the degradation rate of the substrates, which include λ phage replication protein λO, the master regulator of stationary phase RpoS, and SsrA-tagged proteins. TF interacts through multiple sites with ClpX in a highly dynamic fashion to promote protein degradation.

Song *et al.* (Song *et al.* 2020) report an important role of HSP90A at the crossroads between NANOG-TCL1A and multi-aggressive properties of immune-edited tumor cells by identifying chaperone as a NANOG transcriptional target. They found that the HSP90A-TCL1A-AKT pathway ignited by NANOG is a central molecular axis and a potential target for immune-refractory tumors. Girstmair *et al.* (Girstmair *et al.* 2019) studied why *S. cerevisiae* expresses two isoforms of chaperon - Hsp90 (Hsc82, Hsp82). These isoforms differ in their stability even though their sequence identity is very high. They

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

found that despite the differences in the N-terminal ATP-binding domain, the client interactomes are largely identical, but isoform-specific interactors exist both under physiological and heat shock conditions. Wang *et al.* (Wang *et al.* 2019) identified DDO-5936 as a small-molecule inhibitor of the Hsp90-Cdc37 PPI in colorectal cancer. DDO-5936 disrupted the Hsp90-Cdc37 PPI *in vitro* and *in vivo* via binding to a previously unknown site on Hsp90. This leads to the selective down-regulation of Hsp90 kinase clients.

Lim *et al.* (Lim *et al.* 2020) studied the regulation of a tumorigenic factor - AIMP2 lacking exon2 (AIMP2-DX2) by interaction with heat shock protein 70 (HSP70). They presented HSP70 as a critical determinant for the level of AIMP2-DX2. Using X-ray crystallography and NMR analysis, they showed that HSP70 recognizes the amino (N)-flexible terminal region and the glutathione S-transferase domain of AIMP2-DX2 via its substrate-binding domain and blocks the Siah1-dependent ubiquitination of AIMP2-DX2. Hartl group (Imamoglu *et al.* 2020) studied the ATP-dependent HSP70 chaperones (DnaK in E. coli) mediated protein folding in cooperation with J proteins and nucleotide exchange factors (E. coli DnaJ and GrpE). The authors showed that DnaK/DnaJ/GrpE accelerates the folding of the multi-domain protein 20-fold over the rate of folding in the absence of aggregation. DnaK binding expands the misfolded region and thereby resolves the kinetically-trapped intermediates, with folding occurring upon GrpE-mediated release. Mogk and Bukau (Mogk and Bukau 2017) reviewed the molecular mechanisms and working principles of the Hsp70 network. They described how the Hsp70 chaperone system controls diverse cellular functions. This knowledge can lead to the development of chemical compounds that modulate disease-related Hsp70 activities. Ho *et al.* (Ho *et al.* 2019) determined the role of sequestration within the proteostasis network in *S. cerevisiae*. They found that the Heat shock protein 42 and Btn2 sequestrates are intertwined with the refolding activity of the Hsp70 system. Btn2 has chaperone and sequestrate activity and shares features with small heat shock proteins, and during stress conditions, it interacts with the Hsp70 co-chaperone Sis1. Shin *et al.* (Shin *et al.* 2021) studied the LONP1, an AAA+ protease of the mitochondrial matrix. Their results show that LONP1 is a critical factor in the mitochondrial Hsp70 (mtHsp70) pathway. They identified that *in vitro* LONP1 has an intrinsic chaperone-like activity and collaborates with mtHSP70 to stabilize a folding intermediate.

Santagata group (Gaglia *et al.* 2020) studied a heat-shock factor 1 (HSF1), a transcriptional regulator of chaperones. They reported that HSF1 drives chaperone expression during stress but also accumulates separately in nuclear stress bodies called foci. At the same time, they observed HSF1 foci in human tumors and found that their presence inversely correlated with chaperone expression. Tang *et al.* (Tang *et al.* 2020) reported that AKT activates HSF1 via phosphorylation of Ser230. HSF1 physically

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

neutralizes soluble amyloid oligomers (AOs), and beyond amyloidogenesis it shields Hsp60 from direct assault by AOs.

Picard group (Bhattacharya *et al.* 2020) characterized the eucaryote-specific function of the co-chaperone Hop/Stip1/Sti1. They found that eucaryotic cells' deletion of the Hop/Sti1 gene display reduced proteasome activity. These cells are more proficient at preventing protein aggregation and promoting protein refolding. Hartl group (Klaips *et al.* 2020) identified Sis1, an essential Hsp40 co-chaperone of Hsp70, as a critical sensor of proteotoxic stress. Sis1 prevented the formation of dense polyQ inclusions and directed soluble polyQ oligomers towards the formation of permeable condensates. Coelho *et al.* (Coelho *et al.* 2019) showed how multiple chaperones detect and remedy defected membrane protein connexin 32 (Cx32) with single point mutation. ER-membrane complex (EMC) aids in membrane integration of low-hydrophobicity transmembrane segments. When they fail to integrate, these are recognized by the ER-lumenal chaperone immunoglobulin binding protein (BiP). Chitwood and Hegde (Chitwood and Hegde 2020) identified the PAT complex, an intermembrane chaperone that protects TMDs during assembly to minimize misfolding of multi-spanning membrane proteins. This complex connects nascent TMDs that contain unshielded hydrophilic side chains within the lipid bilayer, and it removes concomitant with substrate folding. Eftekharzadeh *et al.* (Eftekharzadeh *et al.* 2019) showed that molecular chaperones Hsp40 and Hsp70 recognize a region of the androgen receptor (AR) N-terminal domain (NTD), that interacts with the AR ligand-binding domain upon activation. They also found that, while the free NTD oligomerizes, binding to Hsp70 increases AR solubility. Faust *et al.* (Faust *et al.* 2020) showed that J-domain proteins (JDPs) could also differ in their interactions with Hsp70 chaperones. They found that the major class B JDPs are regulated by an autoinhibitory mechanism that is not present in other classes. In DNAJB1, the Hsp70-binding sites are intrinsically blocked by an adjacent glycine (Gly)-phenylalanine-rich region. This inhibition can be released upon the interaction of a second site on DNAJB1 with the HSP70 C-terminal tail. Jiang *et al.* (Jiang, Rossi, and Kalodimos 2019) studied the mechanism of binding of the co-chaperone Hsp40. They found that Hsp40 engages the protein client in a highly dynamic fashion using a multivalent binding mechanism that alters the folding properties of the client. The number of binding sites differs in different Hsp40 family members. Hsp70 binding to Hsp40 displaces the unfolded client.

Mas *et al.* (Mas *et al.* 2020) characterized the *E. coli* chaperone Skp monomer-trimer transition. They found that the monomeric state is intrinsically disordered, and the formation of the oligomerization interface initiates the folding of the α-helical coiled-coil arms resulting in the formation of active trimeric Skp. Native client proteins contact all three Skp subunits simultaneously, and their binding shifts the Skp population toward the active trimer. Riezman group (Hannich *et al.* 2019)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

showed that anoxia-associated injury causes accumulation of the non-canonical sphingolipid 1-deoxydihydroceramide (DoxDHCer). Anoxia causes an imbalance between Ser and Ala-derived 1-deoxysphinganine. 1-Deoxysphinganine is incorporated into DoxDHCer, which impairs actin folding via the cytosolic chaperonin TRiC. Prevention of DoxDHCer accumulation resulted in decreased anoxia-induced injury.

4.2 Folding and stability of membrane proteins

Predicting functional effects of missense variants in voltage-gated sodium and calcium channels, Heyne and co-researchers (Heyne *et al.* 2020), inferred loss of function (LOF) (n = 518) and gain of function (GOF) (n = 309) likely pathogenic variants from the disease phenotypes of variant carriers based on the known gene-disease mechanisms of 19 different diseases. By training a machine learning model on sequence- and structure-based features prediction LOF or GOF effects [Receiver operating characteristic, ROC = 0.85] of likely pathogenic missense variants. LOF versus GOF prediction corresponded to molecular LOF versus GOF effects for 87 functionally tested variants in SCN1/2/8A and CACNA1I (ROC = 0.73) and was validated in exome-wide data from 21,703 cases and 128,957 controls.

Domain insertion permissibility-guided engineering of allostery in ion channels (Coyote-Maestas *et al.* 2019) found that permissibility is best explained by dynamic protein properties, such as conformational flexibility. Data and the well-established link between protein dynamics and allostery led to propose that differential permissibility is a metric of latent allosteric capacity in Kir2.1. In supporting this notion, inserting light-switchable domains into sites with predicted latent allosteric capacity renders Kir2.1 activity sensitive to light.

Sequential activation of stromal interaction molecule 1 (STIM1) links Ca^{2+} with luminal domain unfolding (Schober *et al.* 2019) MD simulations with live-cell recordings and determined the sequential Ca2+-dependent conformations of the luminal STIM1 domain upon activation. The authors identified the residues within the canonical and noncanonical EF-hand domains that can bind to multiple Ca2+ ions. A model of STIM1 Ca^{2+} binding is presented and this model refined the currently known initial steps of STIM1 activation on a molecular level.

A thermostable closed SARS-CoV-2 spike protein trimer (Xiong *et al.* 2020) designed mutations in S that allow the production of thermostable, disulfide-bonded S-protein trimers that are trapped in the closed, prefusion state. The authors of the paper demonstrated that the designed, thermostable

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

closed S trimer can be used in serological assays. This protein has potential applications as a reagent for serology, virology, and as immunogen.

Identifying immunologically-vulnerable regions of the hepatitis C virus (HCV) E2 glycoprotein and broadly neutralizing antibodies that target them (Quadeer, Louie, and McKay 2019) predictive *in silico* evolutionary model for E2 that identifies one such region, a specific antigenic domain, making it an attractive target for a robust antibody response. Providing a framework for identifying vulnerable regions of E2 and assessing the potency of particular antibodies results can aid the rational design of an effective prophylactic HCV vaccine.

The structure and mechanism of bactericidal mammalian perforin-2, an ancient agent of innate immunity (Ni *et al.* 2020), reveal the mechanism of perforin-2 activation and activity using atomic structures of pre-pore and pore assemblies, high-speed atomic force microscopy, and functional assays. Perforin-2 forms a pre-pore assembly in which its pore-forming domain points in the opposite direction to its membrane-targeting domain. Acidification then triggers pore formation via a 180° conformational change. This novel and unexpected mechanism prevents premature bactericidal attacks and may have played a key role in the evolution of all perforin family proteins.

The Cystic fibrosis transmembrane conductance regulator (CFTR) trafficking mutations disrupt cotranslational protein folding by targeting biosynthetic intermediates (Shishido *et al.* 2020) show that inherited, disease-causing mutations are located within the first nucleotide-binding domain (NBD1) of CFTR have distinct effects on nascent polypeptides. The observed folding defect is highly dependent on nascent chain length as well as its attachment to the ribosome. Restoration of the NBD1 cotranslational folding defect by second-site suppressor mutations also partially restores folding of full-length CFTR. Nascent folding intermediates can play an essential role in disease pathogenesis and thus provide potential targets for pharmacological correction.

Domain-interface dynamics of CFTR revealed by stabilizing nanobodies (Sigoillot *et al.* 2019) develop nanobodies targeting NBD1 of human CFTR and demonstrate their ability to stabilize both isolated NBD1 and full-length protein. Crystal structures of NBD1-nanobody complexes provide an atomic description of the epitopes and reveal the molecular basis for stabilization. Data uncover a conformation of CFTR involving detachment of NBD1 from the TMD, which contrasts with the compact assembly observed in cryo-EM structures. This interface rearrangement is likely to have major relevance for cystic fibrosis pathogenesis but also for the normal function of CFTR and other ATP-binding cassette proteins.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

A posttranslational modification code for CFTR maturation is altered in cystic fibrosis (Pankow, Bamberger, and Yates 2019). Proper maturation of CFTR is dependent on cross-talk between phosphorylation and methylation events in the regulatory insertion element of the protein. Manipulating these posttranslational modifications prevented the maturation of WT CFTR and instead induced its degradation by ER quality control systems findings identify a minimal quantitative and qualitative PTM code for CFTR maturation that distinguishes correctly folded from misfolded CFTR.

X-ray free-electron laser (XFEL) structures of the human melatonin 2 (MT2) receptor revealed the basis of subtype selectivity. (Johansson *et al.* 2019) reported XFEL structures of the human MT2 receptor in complex with the agonists 2-phenylmelatonin (2-PMT) and ramelteon at resolutions of 2.8 Å and 3.3 Å, respectively, along with two structures of function-related mutants: H2085.46A and N862.50D, obtained in complex with 2-PMT. Functional and kinetic data supported a prominent role for intramembrane ligand entry in both receptors and suggested that there might also be an extracellular entry path in MT2. Findings contributed to a molecular understanding of melatonin receptor subtype selectivity and ligand access modes, which are essential for designing highly selective melatonin tool compounds and therapeutic agents.

Computational design of transmembrane pores (Xu *et al.* 2020) produced artificial pores formed by two concentric rings of α-helices that are stable and monodisperse in both their water-soluble and their transmembrane forms. The ability to produce structurally and functionally well-defined transmembrane pores opens the door to creating channels and pores for a wide variety of applications.

Spatial and temporal alterations in protein structure by epidermal growth factor (EGF) regulate cryptic Cys oxidation (Behring *et al.* 2020) performed time-resolved quantification of the EGF-dependent oxidation of 4200 cys sites in A431 cells. 51% of cys were statistically significantly oxidized by EGF stimulation. EGF induced three distinct spatiotemporal patterns of cys oxidation in functionally organized protein networks, consistent with the spatial confinement model. Unexpectedly, protein crystal structure analysis and MD simulations indicated widespread redox regulation of cryptic Cys residues that are solvent-exposed only upon changes in protein conformation.

Programmed spatial organization of biomacromolecules into discrete, coacervate-based protocells (Altenburg *et al.* 2020) describe the development of a discrete, membrane-bound coacervate-based protocellular platform that utilizes the well-known binding motif between Ni^{2+}-nitrilotriacetic acid and His-tagged proteins to exercise a high level of control over the loading of biologically relevant macromolecules. The platform can accrete proteins in a controlled, efficient, and

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

benign manner, culminating in the enhancement of an encapsulated two-enzyme cascade and protease-mediated cargo secretion, highlighting the potency of this methodology. This approach expands the protocellular toolbox and potentially opens the doors to the development of the next generation of complex yet well-regulated synthetic cells.

In structure-function guided modeling of chemokine-GPCR specificity for the chemokine X-C motif chemokine ligand 1 (XCL1) and its receptor XCR1, Fox *et al.* (Fox *et al.* 2019) applied hybrid approach that combined structure-function data and Rosetta modeling to describe key contacts within a chemokine-GPCR interface. They found that the extreme amino-terminal residues of the chemokine XCL1 (Val1, Gly2, Ser3, and Glu4) contribute a large fraction of the binding energy to its receptor XCR1, whereas residues near the disulfide bond-forming residue Cys11 modulates XCR1 activation. Computational analysis of XCL1-XCR1 interactions revealed functional contacts involving Glu4 of XCL1 and Tyr117 and arginine (Arg)273 of XCR1. Findings demonstrate the utility of a hybrid approach, using biological data and homology modeling to study chemokine-GPCR interactions.

Conformational dynamics of a G protein-coupled receptor helix 8 in lipid membranes is examined in detail (Dijkman *et al.* 2020). In a minority of GPCR structures solved to date, C-terminal amphipathic helix 8 (H8) either is absent or adopts an unusual conformation. The controversial existence of H8 in the class A GPCR neurotensin receptor 1 has been examined for the nonthermostabilized receptor in a functionally supporting membrane environment using electron paramagnetic resonance, MD simulations, and circular dichroism.

Computational design of GPCR allosteric signal transductions has been reported (Chen, Keri, and Barth 2020). They designed 36 dopamine receptor D2 (D2) variants, whose constitutive and ligand-induced signaling agreed well with predictions, repurposed the D2 receptor into a serotonin biosensor, and predicted the signaling effects of more than 100 known GPCR mutations. Results reveal the existence of distinct classes of allosteric microswitches and pathways that define an unforeseen molecular mechanism of regulation and evolution of GPCR signaling. The approach enables the rational design of allosteric receptors with enhanced stability and function to facilitate structural characterization and reprogram cellular signaling in synthetic biology and cell engineering applications.

The structural basis of ligand recognition and self-activation of orphan GPCR 52 (GPR52) has been demonstrated (Lin *et al.* 2020). The authors presented the high-resolution structures of human GPR52 in three states: a ligand-free state, a Gs-coupled self-activation state, and a potential allosteric ligand-bound state. Structures reveal that extracellular loop 2 occupies the orthosteric binding pocket and operates as a built-in agonist, conferring an intrinsically high level of basal activity to GPR52. The

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

reported structures can improve the understanding of other self-activated GPCRs and enable the identification of endogenous ligands, and can guide drug discovery efforts that target GPR52.

Structural underpinnings of resistance to inhibitors of cholinesterase-8A (Ric8A) function as a G-protein α-subunit chaperone and guanine-nucleotide exchange factor has been reported in the paper (Srivastava, Gakhar, and Artemyev 2019). The study described two crystal structures of Ric8A, one in the apo form, and the other in complex with a tagged C-terminal fragment of Gα. These structures reveal two main domains: an armadillo-fold core and a flexible C-terminal tail. Additionally, the Gα C-terminus binds to a highly-conserved patch on the concave surface of the Ric8A armadillo-domain, with selectivity determinants residing in the Gα sequence. Ric8A C-terminal tail is critical for its stability and function. The above-mentioned study of Srivastava *et al.* lays the groundwork for understanding Ric8A function at the molecular level.

4.3Amyloids

Schymkowitz *et al.* (Michiels *et al.* 2020) studied the interaction between human amyloids with viruses and intervention with vital replication. The authors engineered virus-specific amyloids against influenza A and Zika proteins. Each amyloid shares a homologous aggregation-prone fragment with a specific viral target protein. For influenza, they demonstrated that designer amyloid against PB2 accumulates in influenza A-infected tissue *in vivo*. Brockwell group (Ebo *et al.* 2020) used an assay in the periplasm of *E. coli* linking aggregation directly to antibiotic resistance acts to detect the innate aggregation of antibody fragments. They demonstrated the generation of aggregation-resistant scFv sequences when reformatted as immunoglobulins G. Li *et al.* (Li, Li, *et al.* 2020) reported *Escherichia coli* biofilm-inspired protein nanofiber coatings that exhibit substrate independence, resistance to organic solvents, and programmable functionalities. The intrinsic surface adherence of CsgA amyloid proteins facilitates forming nanofiber coatings on any surface with varied compositions, sizes, shapes, and structures. Based on their genetically engineerable functionality, nanofiber coatings can participate in diverse protein conjugations, deoxyribonucleic acid (DNA) binding, thus enabling a variety of applications, including electronic devices, enzyme immobilization, and microfluidic bacterial sensors.

Nirmalraj *et al.* (Nirmalraj *et al.* 2020) demonstrated the interface between graphene and pure water as an ideal platform for resolving the size, shape, and morphology of Aβ-40 and Aβ-42 peptide assemblies with single-particle resolution. Differences in Aβ-40 and Aβ-42 networks were resolved, wherein only Aβ-42 fibrils were aligned through lateral interactions over micrometer-scale lengths.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Yee *et al.* (Yee *et al.* 2019) proposed a mechanism whereby human transthyretin (TTR) can form amyloid fibrils via a parallel equilibrium of partially unfolded species that proceeds in favor of the amyloidogenic forms of TTR. It is suggested that unfolding events within the TTR monomer originate at the C-D loop of the protein and that destabilizing mutations in this region enhance the rate of TTR fibrillation. Wills-Karp (Smole *et al.* 2020) reported that mite group 13 allergens of the fatty acid-binding protein (FABP) family are sensed by an acute-phase protein, serum amyloid A1 (SAA1), that promotes pulmonary type 2 immunity. SAA1 interacted directly with allergenic mite FABPs. The interaction between mite FABPs and SAA1 activated the SAA1-binding receptor, formyl peptide receptor 2, which drove the epithelial release of the type-2-promoting cytokine interleukin (IL)-33 in a SAA1-dependent manner. Yu *et al.* (Yu, Zhang, *et al.* 2020) identified the oligomerization states of an amyloid probe thioflavin T (ThT) on hIAPP8-37 assembly. They demonstrated that both adhesive interactions between ThT and the protein substrate and cohesive interactions among ThT molecules govern the oligomerization state of the bounded ThT. Petranovic *et al.* (Chen, Ji, *et al.* 2020) reported a genome-wide synthetic genetic interaction array to identify toxicity modifiers of Aβ42, using yeast as the model organism. They found that flavin mononucleotide (FMN) 1, a gene encoding riboflavin kinase and its product FMN, reduce Aβ42 toxicity. They showed the effects of FMN supplementation to include reducing misfolded protein load and increasing resistance to oxidative stress. The Hong group (Gelenter *et al.* 2019) determined the atomic-resolution structure of fibrils of synthetic human glucagon grown at low pH. Data indicates the coexistence of two β-strand conformations, which have distinct water accessibilities and intermolecular contacts, indicating that they alternate and hydrogen bond in an antiparallel fashion along the fibril axis. Two antiparallel β-sheets assemble with symmetric homodimer cross-sections. Swuec *et al.* (Swuec *et al.* 2019) reported the 4.0 Å resolution cryo-electron microscopy map and molecular model of amyloid fibrils extracted from the heart of an AL amyloidosis patient with amyloid cardiomyopathy. The helical fibrils are composed of a single protofilament showing 4.9 Å stacking and cross-β architecture. Two polypeptide stretches (77 residues) from the light-chain (LC) variable domain fit the fibril density. Residues stabilizing the fibril core are conserved through different cardiotoxic, highlighting structural motifs that may be common to misfolding-prone LCs. Li *et al.* (Sun *et al.* 2020) determined the amyloid fibril structure formed by the heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) LC domain. The structure reveals that the nuclear localization sequence of hnRNPA1 (PY-NLS), which is initially known to mediate the nucleo-cytoplasmic transport of hnRNPA1 through binding with karyopherin-β2 (Kapβ2), represents the major component of the fibril core. The residues that contribute to the binding of PY-NLS with Kapβ2 also exert key molecular interactions to stabilize the fibril structure. Röder *et al.* (Roder *et al.* 2019) presented the atomic model of the phosphatidyl-inositol-3-kinase amyloid fibril. The fibril is composed of two intertwined

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

protofilaments that create an interface spanning 13 residues from each monomer. The model comprises residues 1-77 out of 86 amino acids in total, with the missing residues located in the highly flexible C-terminus. Radamaker *et al.* (Radamaker *et al.* 2019) reported the structure of a λ1 AL amyloid fibril from an explanted human heart. The fibril core consists of a 91-residue segment presenting an all-beta fold with ten mutagenic changes compared to the germlin. The conformation differs substantially from natively folded light chains - a rotational switch around the intramolecular disulfide bond being the crucial structural rearrangement underlying fibril formation. Liberta *et al.* (Liberta *et al.* 2019) reported the purification and electron cryo-microscopy analysis of amyloid fibrils from a mouse and a human patient with systemic AA amyloidosis. The two fibrils differ in fundamental properties, such as the presence of right-hand or left-hand twisted cross-β sheets and the overall fold of the fibril proteins. Both proteins adopt highly similar β-arch conformations within the N-terminal ca. 21 residues. Schmidt *et al.* (Schmidt *et al.* 2019) presented a cryo-electron microscopy structure of a fibril purified from the tissue of a patient with hereditary Val30Met WT transthyretin amyloidosis. The fibril consists of a single protofilament that is formed from an N-terminal and C-terminal fragment of transthyretin. The Eisenberg group (Cao *et al.* 2019) found the features of TAR DNA-binding protein 43 (TDP-43) fibrils that confer both reversibility and irreversibility by determining structures of two segments reported to be the pathogenic cores of human TDP-43 aggregation. Energetic analysis suggests that the dagger-shaped polymorphs represent irreversible fibril structures, whereas the SegB polymorph may participate in both reversible and irreversible fibrils. The Schröder Group (Roder *et al.* 2020) reconstructed densities of three dominant islet amyloid polypeptide (IAPP) fibril polymorphs. An atomic model reveals two S-shaped, intertwined protofilaments. The segment 21-NNFGAIL-27, essential for IAPP amyloidogenicity, forms the protofilament interface together with Tyr37 and the aminated C terminus. The S-fold resembles polymorphs of Alzheimer´s disease-associated amyloid-β fibrils, which might account for the epidemiological link between type 2 diabetes and Alzheimer´s disease and reports on IAPP-Aβ cross-seeding *in vivo*.

4.4 Degradation of proteins: role of stability and folding

Wu *et al.* (Wu *et al.* 2020) reported a structure of the active Hrd1 complex, determined by cryo-electron microscopy analysis of two subcomplexes. Hrd3 and Yos9 jointly create a luminal binding site that recognizes glycosylated substrates. These structures, along with crosslinking and MD simulation results, suggest how a polypeptide loop of an ER-associated degradation (ERAD)-L substrate moves through the ER membrane. The Bachovchin group (Chui *et al.* 2019) discovered that lethal factor

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

induces cell death via the N-end rule proteasomal degradation pathway. Lethal factor directly cleaves NLRP1B, yielding the N-end rule-mediated degradation of the NLRP1B N terminus and freeing the NLRP1B C terminus to activate caspase-1. DPP8/9 inhibitors also induce proteasomal degradation of the NLRP1B N terminus but not via the N-end rule pathway. N-terminal degradation is the common activation mechanism of this innate immune sensor. Mueller *et al.* (Mueller *et al.* 2021) showed that suppression of the N-alpha-acetyltransferase NatA turns these cryptic IBM-like sequences into a very efficient inhibitor of apoptosis protein (IAP) binders in cell lysates and in vitro and ultimately triggers cellular apoptosis. Amino-terminal acetylation of IBM-like motifs in NatA substrates shields them from IAPs. This previously unrecognized relationship suggests that amino-terminal acetylation is generally protective against protein degradation in human cells. It also identifies IAPs as agents of a general quality control mechanism targeting unacetylated rogues in metazoans. The Xu group (Chen *et al.* 2021) uncovered the molecular mechanism of Arg/C-degron recognition by solving a subset of structures of FEM1 proteins in complex with Arg/C-degron-bearing substrates. Structural research, complemented by binding assays and global protein stability analyses, demonstrate that FEM1A/C and FEM1B selectively target distinct classes of Arg/C-degrons. The study also provides valuable information for developing chemical probes for selectively regulating proteostasis. Yan *et al.* (Yan *et al.* 2021) present crystal structures of FEM1C in complex with Arg/C-degron and show that FEM1C utilizes a semi-open binding pocket to capture the C-terminal arg and that the extreme C-terminal arg is the major structural determinant in recognition by FEM1C.with biochemical and mutagenesis studies, provide a framework for understanding molecular recognition of the Arg/C-degron by the FEM family of proteins.

Ribosome-associated protein quality control (RQC) is now known to have a role on the surfaces of the ER and mitochondria. Joazeiro *et al.* (Joazeiro 2019) described the knowledge on RQC mechanisms, highlighting key features of Ltn1/listerin action that provide a paradigm for understanding how E3 ligases operate in protein quality control in general review discuss how defects in this pathway may compromise cellular function and lead to disease. Crews *et al.* (Schapira *et al.* 2019) described targeted protein degradation in chemical biology and drug discovery. They systematically reviewed the expression profile, domain architecture, and chemical tractability of human E3 ligases to expand the toolbox for proteolysis targeting chimera discovery.

The Sonja Lorenz group (Liess *et al.* 2020) uncovered a dimeric state of ubiquitin-conjugated E2 S enzyme (UBE2S) that confers autoinhibition by blocking a catalytically critical ubiquitin-binding site. Dimerization is stimulated by the lysine-rich carboxyl-terminal extension of UBE2S that is also required to recruit this E2 to the anaphase-promoting complex/cyclosome and is autoubiquitination

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

as substrate abundance becomes limiting. Found that dimerization-deficient UBE2S turned over more rapidly in cells and did not promote mitotic slippage during prolonged drug-induced mitotic arrest. Data illustrate how the use of mutually exclusive macromolecular interfaces enables modulation of both the activities and the abundance of E2s in cells to facilitate precise ubiquitin signaling.

Kopp *et al.* (Kopp *et al.* 2019) reconstituted components of human unfolded protein response (UPR), ER stress, and BiP chaperone systems discover that the interaction of BiP with the luminal domains of UPR proteins inositol-requiring enzyme 1 (IRE1) and protein kinase R-like endoplasmic reticulum kinase (PERK) switch BiP from its chaperone cycle into an ER stress sensor cycle by preventing the binding of its co-chaperones, with loss of adenosine triphosphatase stimulation Furthermore, misfolded protein-dependent dissociation of BiP from IRE1 is primed by ATP but not adenosine diphosphate. Elucidation of the previously unidentified mechanical cycle of the BiP function explains its ability to act as an Hsp70 chaperone and ER stress sensor.

The Aebersold group (Shao *et al.* 2019) investigate the degradation of messenger RNA (mRNA) and protein in 68 pairs of adjacent prostate tissue samples using RNA sequencing and sequential window acquisition of all theoretical mass spectra MS to develop a numerical score, the Proteome Integrity Number (PIN), that measures the degree of protein degradation. Results indicate that protein degradation only affects 5.9% of the samples tested and shows negligible correlation with mRNA degradation in the adjacent samples. Independent analyses confirm findings on additional clinical sample cohorts and across different mass spectrometric methods Overall, the data show that most samples tested are not compromised by protein degradation and establish the PIN score as a generic and accurate indicator of sample quality for proteomic analyses.

Li *et al.* (Li, Sun, *et al.* 2020) reported that SE binds 20S core proteasome α subunit G1 (PAG1) among other components and is accumulated in their mutants. Purified PAG1-containing 20S proteasome degrades recombinant SE via an ATP- and the ubiquitin-independent manner *in vitro*. *PAG1* is a positive regulator for *SE* in vivo, as *pag1* shows comparable molecular and/or developmental defects relative to *se*. SE is poorly assembled into macromolecular complexes study provides insight into how the 20S proteasome regulates RNA metabolism through controlling its key factor in eukaryotes.

The Kaufman group (Hetz, Zhang, and Kaufman 2020) described ER proteostasis surveillance as mediated by the UPR. This signal transduction pathway senses the fidelity of protein folding in the ER lumen. Recent advances in understanding the regulation of UPR signaling and its implications in the

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

pathophysiology of disease might open new therapeutic avenues. Fernando Aprile-Garcia *et al.* (Aprile-Garcia *et al.* 2019) enhanced recruitment of negative transcription elongation factors to gene promoters in human cell lines induces stress-induced transcriptional attenuation. The chemical inhibitor screen showed that active translation and protein ubiquitination are required for the response. Proteins translated during heat shock are subjected to ubiquitination, and p38 kinase signaling connects cytosolic translation with gene downregulation.

Liu *et al.* (Liu, Guan, *et al.* 2020) showed that ubiquitin fold modifier 1 can covalently modify p53 and that this modification stabilizes p53 by antagonizing its ubiquitination and proteasome degradation. Clinically, UFL1 and DDRGK1 expression are downregulated and positively correlated with levels of p53 in a high percentage of renal cell carcinomas. Results identify UFMylation as a crucial post-translational modification for maintaining p53 stability and tumor-suppressive function and point to UFMylation as a promising therapeutic target in cancer.

Mena Abdel-Nour *et al.* (Abdel-Nour *et al.* 2019) found that the eukaryotic initiation factor 2α (eIF2α) kinase heme-regulated inhibitor (HRI) controls nucleotide-binding oligomerization domain containing 1 signalosome folding and activation through a process requiring eIF2α, the activating transcription factor 4, and the heat shock protein heat shock protein family B Member 8 (HSPB8). The HRI/eIF2α signaling axis was also essential for signaling downstream of the innate immune mediator's nucleotide-binding oligomerization domain containing 2, mitochondrial antiviral-signaling protein, and TIR-domain-containing adapter-inducing interferon-β but dispensable for pathways dependent on myeloid differentiation primary response 88 or stimulator of interferon genes. Propose that HRI, eIF2α, and HSPB8 define a novel cytosolic UPR essential for optimal innate immune signaling by large molecular platforms, functionally homologous to the PERK/eIF2α/HSPA5 axis of the ER UPR.

The Hartl group reviewed (Hipp, Kasturi, and Hartl 2019) stresses that lead to the decline of proteostasis network capacity and proteome integrity. The resulting accumulation of misfolded and aggregated proteins, particularly postmitotic cell types such as neurons, manifests in disease. Recent analyses of proteome-wide changes during aging inform strategies to improve proteostasis. The possibilities of pharmacological augmentation of the capacity of proteostasis networks hold great promise for delaying the onset of age-related pathologies associated with proteome deterioration and for extending health span. The Susan Marqusee group (Carroll *et al.* 2020) developed a generalizable strategy to produce isopeptide-linked ubiquitin within structured protein regions. The effects on the energy landscape vary from negligible to dramatic, depending on the protein and site of ubiquitination. Ubiquitination at sensitive areas destabilizes the native structure and increases the rate of

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

proteasomal degradation. In well-folded proteins, ubiquitination can even induce the requisite unstructured regions needed for proteasomal engagement. Xhani *et al.* (Xhani *et al.* 2020) reported that the structured DNA binding domain (DBD) of purine-rich box-1 regulates gene expression via antagonistic dimeric states that are reciprocally controlled by cognate DNA on the one hand and by its proximal anionic IDR on the other. The two conformers are mediated by distinct regions of the DBD without structured contributions from the tethered IDRs. Unlike DNA-bound complexes, the unbound dimer is markedly destabilized. These results suggest a previously unidentified, nonstructural role for charged IDRs in conformational control by mitigating electrostatic penalties that mask the interactions of highly cationic DBDs.

4.5Protein Switches

Carrasco-López *et al.* (Carrasco-Lopez *et al.* 2020) demonstrated proof-of-principle, i.e., the development of a light-controlled monobody (OptoMB) that works in vitro and cells and whose affinity for its SH2-domain target exhibits a 330-fold shift in binding affinity upon illumination. They demonstrate that αSH2-OptoMB can be used to purify SH2-tagged proteins directly from *E. coli* extract, achieving 99.8% purity and over 40% yield in a single purification step. OptoMBs have the potential to find new applications as light-switchable binders of untagged proteins with the temporal and spatial precision afforded by light. Favretto *et al.* (Favretto *et al.* 2020) provided atomic insight into a tug-of-war between cis/trans isomerization and molecular chaperone activity. They showed that cis/trans isomerization outpowers the holding activity of cyclophilin A. Removal of the Pro isomerization barrier through posttranslational truncation of α-synuclein reverses the action of the Pro isomerase and turns it into a potent molecular chaperone that inhibits protein misfolding. The data reveal a conserved mechanism of the dual functionality in cis/trans isomerases and define its molecular determinants acting on intrinsically disordered proteins. Escobedo *et al.* (Escobedo *et al.* 2019) studied the structural basis of the association between tract length, transcriptional activity, and disease how the conformation of the polyQ tract of the AR, associated with spinobulbar muscular atrophy (SBMA), depends on its length. Report that this sequence folds into a helical structure stabilized by unconventional hydrogen bonds between glutamine side chains and main-chain carbonyl groups. Its helicity directly correlates with tract length. Findings suggest a plausible rationale for the association between polyQ tract length and AR transcriptional activity and have implications for establishing the mechanistic basis of SBMA. Kumar *et al.* (Kumar *et al.* 2020) determine the atomic resolution structures of dimeric, tetrameric, and pentameric immunoglobulin A (IgA)-fragment crystallizable

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

region linked by the joining chain and complex with the secretory component of the polymeric immunoglobulin receptor. This framework will inform the design of future IgA-based therapeutics.

Toyama *et al.* (Toyama *et al.* 2019) investigated the activation mechanism of the N92I mutant with NMR spectroscopy; the mutation does not markedly perturb the static structure of the guanosine diphosphate (GDP) binding site, but the overall conformational stability decreases in the N92I mutant, which then facilitates GDP dissociation by lowering the activation energy for the dissociation reaction. Based on these results, the activation mechanism of the N92I mutant, in which the decreased conformational stability, plays an essential role in its activation process.

The Lomas group (Faull *et al.* 2020) isolated polymers from the liver tissue of Z α1-antitrypsin homozygotes (E342K) who have undergone transplantation, labeled them using a Fab fragment, and performed a single-particle analysis of negative-stain electron micrographs. The data show structural equivalence between heat-induced and ex vivo polymers and that the intersubunit linkage is best explained by a carboxyl-terminal domain swap between molecules of α1-antitrypsin.

Okumura *et al.* (Okumura *et al.* 2019) present mechanisms of action of protein disulfide isomerase (PDI)—the most versatile disulfide-introducing enzyme in the ER—during the catalysis of oxidative protein folding. Single-molecule analysis by high-speed atomic force microscopy revealed that oxidized PDI is in rapid equilibrium between open and closed conformations. In contrast, reduced PDI is maintained in the closed state. To effectively guide proper oxidative protein folding, PDI regulates conformational dynamics and oligomeric conditions by its redox state and the configurations or folding states of substrates.

SthK exists in two forms: trans Pro300 SthK with high ligand binding affinity and fast activation and cis Pro300 SthK with low affinity and slow activation. Schmidpeter *et al.* (Schmidpeter, Rheinberger, and Nimigean 2020) investigate slow cyclic adenosine monophosphate-induced activation in purified SthK channels using stopped-flow assays and mutagenesis enzymatic catalysis and inhibition assays, revealing that the cis/trans conformation of a conserved Pro in the cyclic nucleotide-binding domain determines the activation kinetics of SthK. Results reveal prolyl isomerization as a regulatory mechanism for SthK, and potentially eukaryotic hyperpolarization-activated cyclic nucleotide channels. This mechanism could contribute to electrical rhythmicity in cells.

The Baker group (Langan *et al.* 2019)explored the possibility of designing switchable protein systems *de novo* through the modulation of competing for inter-and intramolecular interactions. They described orthogonal cage–critical systems that function in vitro, in yeast and mammalian cells with up to 40-fold activation of function by key. They designed a static, five-helix ‘cage’ with a single

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement N° 952333

interface that can interact either intramolecularly with a terminal 'latch' helix or intermolecularly with a peptide 'key'. The study opens up new avenues for synthetic biology and cell engineering.

The Volkman group (Dishman *et al.* 2021)studied the evolution of the metamorphic human protein XCL1, which has two distinct folds with different functions, making it an unusual member of the chemokine family, whose members generally adopt one conserved fold with using ancestral reconstruction and NMR discoveries illuminate how one sequence has evolved to encode multiple structures, revealing principles for protein design and engineering. The Munoz group (Campos *et al.* 2019) presented a proof of concept of allosteric assembly in which an engineered fold switch on the protein monomer triggers or blocks assembly. The design is based on the hyper-stable, naturally monomeric protein chymotrypsin inhibitor 2 (CI2), a paradigm of simple two-state folding, and the toroidal arrangement with 6-fold symmetry that it only adopts in crystalline form. They engineered CI2 to enable a switch between the native and an alternate, latent fold that self-assembles onto hexagonal toroidal particles by exposing a favorable inter-monomer interface. Findings unveil a remarkable potential for structural metamorphosis in proteins and demonstrate critical principles for engineering protein-based nanomachinery. Leopold *et al.* (Leopold *et al.* 2019) combined catalytic domains of Trk (tropomyosin receptor kinase) family of receptor tyrosine kinases (RTKs), naturally activated by neurotrophins, with a photosensory core module of photosensory module truncated from *Deinococcus radiodurans* phytochrome to develop opto-kinases, termed Dr-TrkA and Dr-TrkB, reversibly switchable on and off with near-infrared and far-red light. Validated Dr-Trk's ability to reversibly light-control several RTK pathways, calcium levels and demonstrated that their activation triggers canonical Trk signaling. Dr-Trks have several superior characteristics that make them the Opto-kinases of choice for regulation of RTK signaling: high activation range, fast and reversible photoswitching, and multiplexing with visible-light-controllable optogenetic tools.

Zhang *et al.* introduced a new strategy for reversible off-on control of enzyme activity by near-infrared light. Reversibly switch off and on the activities of three enzymes acting on polysaccharides, protein, and plasmid. The enzyme activities are increased by up to 61-fold after laser irradiation. The study provides a method for off-on control of enzyme activity. It is introduced the concept of an allosteric force that can be used to drive chemical reactions - based on the energetics of the four states of the system (cis vs. trans and ligand-bound vs. free). To design a microscopic understanding of this process, Bozovic *et al.* (Bozovic, Jankovic, and Hamm 2020) a protein system has been designed - allosteric communication between the binding and allosteric site can be observed in both directions.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

The development of a light-activated protein phosphatase using the dual-specificity phosphatase 6 (or MKP3) was described (Courtney and Deiters 2019). Through genetic code expansion, MKP3 is placed under optical control via two different approaches: (i) incorporation of a caged cys into the active site for controlling catalytic activity and (ii) incorporation of a caged lysine into the kinase interaction motif for controlling the PPI between the phosphatase and its substrate. Applying the optogenetically controlled MKP3 in conjunction with live-cell reporters, they discover that ERK nuclear translocation is regulated in a graded manner in response to increasing MKP3 activity.

Lee *et al.* (Lee *et al.* 2020) developed a SPOTlight, which is a high-throughput technique to isolate individual yeast or human cells with unique spatiotemporal profiles from heterogeneous populations. It relies on visual imaging phenotypes by microscopy, precise optical tagging of single target cells, and retrieval of tagged cells by fluorescence-activated cell sorting. Screened 3 million cells expressing mutagenesis libraries and identified a bright new variant, mGold, the most photostable yellow fluorescent protein reported to date. The SPOTlight will facilitate its deployment to decipher the rules of life, understand diseases, and engineer new molecules and cells.

Yao *et al.* (Yao *et al.* 2020) used a pair of cellulosomal assembly modules, comprising cohesin and a dockerin from Clostridium acetobutylicum, which interact together following a unique pH-dependent switch between two functional sites rather than on-off states. The two cohesin-binding sites on the dockerin are switched from one to the other at pH 4.8 and 7.5 with a 180° rotation of the bound dockerin. Combined analysis by NMR spectroscopy, crystal structure determination, mutagenesis, and isothermal titration calorimetry elucidates the chemical and structural mechanism of the pH-dependent switching of the binding sites. Gabryelczyk *et al.* (Gabryelczyk *et al.* 2019) applied histidine-rich squid beak proteins (HBPs) as model IDPs to shed light on molecular interactions governing liquid-liquid phase separation (LLPS). Show that LLPS of HBPs is mediated through specific modular repeats. Findings provide guidelines to rationally design pH-responsive peptides with LLPS ability, including bioinspired protocells and intelligent drug-delivery systems.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

5. SELECTED PUBLICATIONS ON PROTEIN STABILITY AND FOLDING

5.1 Regulation of chaperone function by coupled folding and oligomerization

Mas, G., B. M. Burmann, T. Sharpe, B. Claudi, D. Bumann, and S. Hiller. 2020. 'Regulation of chaperone function by coupled folding and oligomerization,' *Sci Adv.* 2020 Oct 21;6(43):eabc5822.

Molecular chaperones are central to the survival of the cell. They are involved in many cellular processes, helping proteins fold, preventing protein aggregation, and reducing cellular stress. The homotrimeric molecular chaperone Skp is an ATP-independent chaperone, which facilities the transport of outer membrane proteins across the periplasm. Its activity is modulated during its functional cycle and is connected with the monomer-trimer transition. The monomeric state of Skp is intrinsically disordered. The formation of the oligomerization interface initiates the folding of the α-helical coiled-coil arms via a unique mechanism, resulting in the formation of active trimeric Skp. Native client proteins contact all three subunits simultaneously, and their binding shifts Skp toward the functional trimer.

For structural characterization of monomeric Skp, mutants that destabilize the oligomerization interface to shift the oligomerization equilibrium toward the monomeric form were designed. The structure of Skp is stabilized by a network of three β-sheets per subunit that together form the trimerization interface (Fig. 5.1.1A). The conserved alanine-103 and alanine-108 were identified because they are located at the oligomerization interface with limited space for their side chains. Replacement of alanine for leucine or arg should destabilize the trimer (Fig. 5.1.1A, B). The mutant V117P was designed, insertion of a Pro residue, which is known, leads to break of secondary structure. Oligomerization state of each of the mutant - Skp(A103L), Skp(A103R), Skp(A108L), Skp(A108R), Skp (V117P) was determined by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) at an elution concentration of ≈ 80 µM. At this concentration, the WT protein is mostly trimeric. The mutant A103L was hardly distinguishable from WT, but other mutants featured increased monomeric fraction in order A103R, A108L, V117P, and A108R (Fig. 5.1.1C). Next, the dose-dependence of the equilibrium between monomer/oligomers was quantified for Skp mutants - A103L, A103R, A108L by NMR spectroscopy and SEC-MALS (Fig. 5.1.1C to E). Skp(WT) followed an equilibrium with $C_{0.5}$ = 1.5 µM. Skp (A103L) showed a trimer-monomer equilibrium that was identical to WT (Fig.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

5.1.1E), Skp (A103R) had $C_{0.5}$ = 7 ± 2 µM and Skp(A108L) had $C_{0.5}$ = 80 ± 20 µM (Fig. 5.1.1D). Two mutants - Skp(A108R) and Skp(V117P) were found to be monomeric at concentrations up to 1 mM (Fig. 5.1.1E).

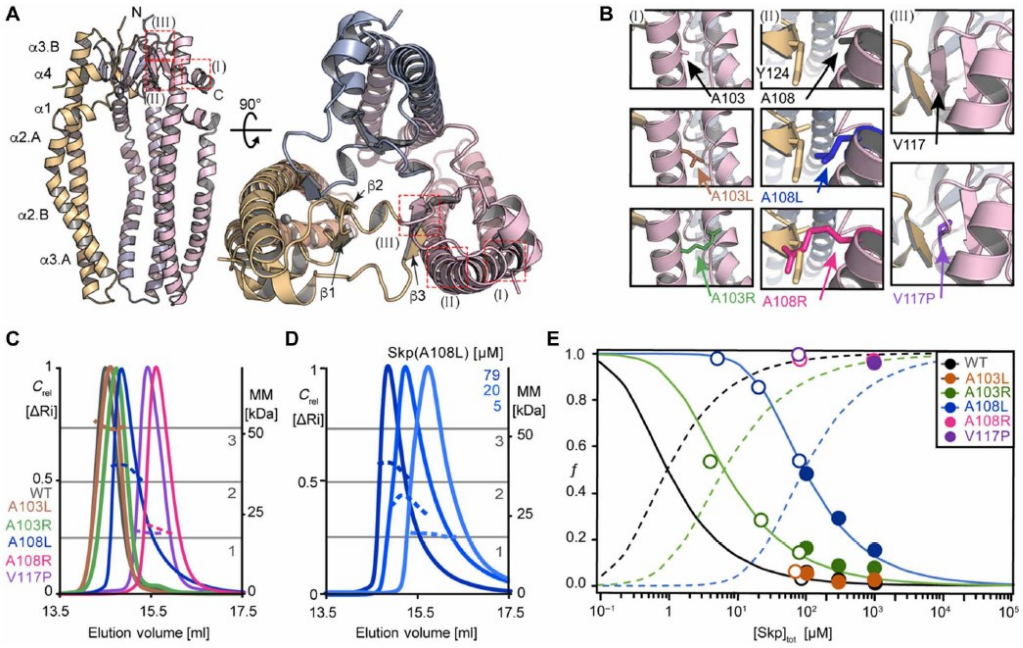

Figure 5.1.1 Identification of Skp mutations that perturb the monomer-trimer equilibrium. (A) Location of the mutation sites [red boxes (I), (II), and (III)], displayed on the Skp crystal structure (Protein Data Bank: 1SG2). Secondary structure elements and termini are indicated. (B) Close-up of the interface between Skp subunits, highlighting the position of the five mutations. See text for details. (C) SEC elution profiles (solid lines, left axis) and MALS apparent molecular mass (MM) (dotted lines, right axis) at elution concentrations of ≃80 µM and a temperature of 25°C. Dark gray, Skp(WT); brown, Skp(A103L); green, Skp(A103R); blue, Skp(A108L); magenta, Skp(A108R); purple, Skp(V117P). Gray horizontal lines indicate the molecular masses of monomers, dimers, and trimers of Skp. (D) Experiment as in (C) for Skp(A108L) as a function of the elution concentrations: 5, 20, and 79 µM. (E) Fractional populations f of monomers in the monomer-trimer equilibrium as a function of total Skp concentration. Experimental data points from NMR and SEC-MALS are indicated by filled and open circles, respectively. These have been fitted by Eq. 4 for mutants A103R and A108L (solid lines). The corresponding fractional populations of the trimeric state, 1 − f, are shown by dashed lines. Note that the concentration of Skp trimers equals one-third of the concentration of Skp molecules in the trimeric state, i.e., [Skp]trimer = 1/3 · (1 − f) · [Skp]tot.

The NMR spectra of the mutant Skp(A108L) with a $C_{0.5}$ = 80 ± 20 µM at 25 °C are entirely overlapping with the monomeric but low-abundant conformation of Skp(WT) (Fig. 5.1.2A). Next, the authors quantified the secondary structure elements, complete sequence-specific resonance assignments of the monomeric state were established and determined backbone ^{13}Cα and ^{13}Cβ secondary CS (Fig. 5.1.2B). Three β-sheets that constitute the oligomerization interface in the trimeric form are in random-coil conformation in the monomeric state. The four α-helices forming the arms of Skp are in a fast conformational exchange between folded and unfolded conformations. The residual helicity can be quantified for each residue by taking the fully denatured form of Skp in 8 M urea and the folded trimer (Fig. 5.1.2B, C). The analysis shows the helices α1, α3.B, and α4, which are closest to the trimerization interface, feature a residual helicity of <20%, the helices α2.A, α2.B, and α3.A located

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

at the tip of the arms display a helical population of 20 to 30% (Fig. 5.1.2C, D). Data demonstrate that the monomeric state of Skp is intrinsically disordered, with some residual helical propensity located at the tip of the arms. In the trimeric structure, the circular-barrel interface, connecting the N- and C-terminal part of the protein, brings helices α2 and α3 close together in space and thus stabilizes secondary structure (Fig. 5.1.2E).

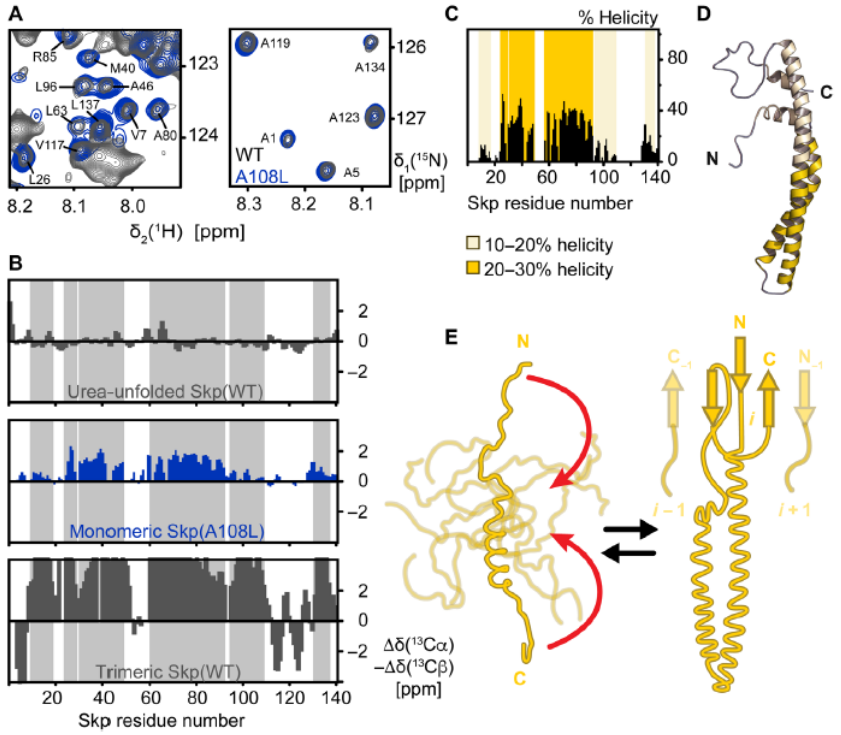

Figure 5.1.2 Monomeric Skp is intrinsically disordered. (A) Sections of 2D [15N,1H]-TROSY spectra of [U-2H,15N] Skp(WT) (dark gray) and Skp(A108L) (blue) at a concentration of 1 mM and 37°C in NMR buffer (20 mM MES, pH 6.5, and 150 mM NaCl). NMR signals of the monomeric state of Skp(WT) are overlaying with the one from Skp(A108L). The assignments of the overlapping NMR signals of the monomeric state are indicated in the panel. (B) Residue-specific secondary backbone chemical shifts of Skp(WT) in 8 M urea solution, Skp(A108L) in its monomeric form, and Skp(WT) in its trimeric form. Positive and negative values indicate α-helical and β-sheet secondary structure elements, respectively. The gray-shaded area indicates the positions of helices in the Skp trimer. (C) Percentage of helical population in the conformational ensemble of the Skp monomer. Helical regions with 10 to 20% helicity or 20 to 30% helicity are highlighted with light or dark yellow, respectively. (D) Structural model of the Skp monomer. On a configuration of Skp with α-helices formed, the degree of residual helical population present in the conformational ensemble is indicated. The large majority of monomeric Skp is disordered. (E) Schematic model of coupled oligomerization and folding mechanism of Skp. Monomeric Skp explores an ensemble of conformations with a low propensity for the formation of the arm α-helices. The formation of the oligomerization interface brings the N and C termini together (red arrows), thus stabilizing the coiled-coil structure of the α-helical arms.

Next, the research was focused on understanding how the equilibrium between folded trimer and a disordered monomer contributes to Skp chaperone activity. As a model client was used the native client protein tOmpA, an eight-stranded TMD of OmpA. The first step was to measure the chaperone activity by quantifying the Skp-bound tOmpA. The activity correlated with the

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

concentration of the trimer for all Skp variants, Skp(A108L) has around 50% of the Skp(WT) activity, and that no chaperone activity was detected for Skp(V117P) and Skp(A108R) (Fig. 5.1.3A, B).

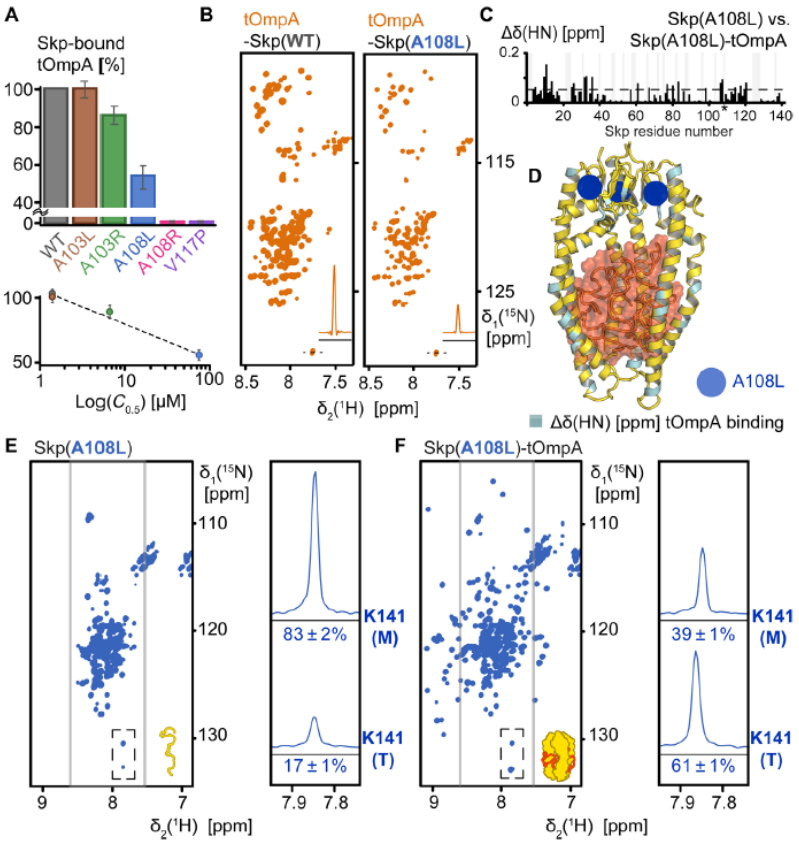

Figure 5.1.3 Client binding promotes Skp oligomer assembly. (A) Holdase activity of Skp variants as determined by the amount of aggregation-prone tOmpA solubilized in equilibrium. Values are normalized to the activity of Skp(WT). Error bars represent the SD of 15 individual signals of tOmpA. (B) 2D [15N,1H]-TROSY fingerprint spectra of [U-2H,15N]-tOmpA bound to unlabeled Skp(WT) or Skp(A108L). Spectra were recorded at a temperature of 37°C in NMR buffer (20 mM MES, pH 6.5, and 150 mM NaCl). A 1D 1H cross section shows the intensity of alanine-176. (C) Combined amide chemical shift differences between [U-2H,15N]-Skp(WT) and [U-2H,15N]-Skp(A108L) with bound unlabeled tOmpA. The magnitude of 2 SDs [0.053 parts per million (ppm)] is indicated by a dashed line. (D) Structural model of Skp(108L) with bound tOmpA. Amide groups with chemical shift changes larger than 2 SDs upon binding of tOmpA to Skp(A108L) are marked in light blue. The position of A108 is indicated by a blue circle. (E and F) 2D [15N,1H]-TROSY fingerprint spectra of [U-2H,15N] Skp(A108L) in the absence (E) and presence (F) of unlabeled tOmpA. Spectra were recorded at 37°C in NMR buffer. The spectral area 7.5 to 8.5 ppm in 1H, corresponding to disordered protein states, is indicated by gray lines. 1D 1H cross sections of lysine-141 in the monomeric (M) and trimeric (T) state of Skp are shown, and the relative fractions are indicated.

Skp(A108L) was selected to characterize the structure and arrangement of the tOmpA-Skp(A108L) complex. The two-dimensional NMR spectra of isotope-labeled tOmpA bound to unlabeled Skp(A103L), Skp(A103R), Skp(A108L), and Skp(WT) are highly similar (Fig. 5.1.3B). This observation indicates that the client conformational ensemble inside the chaperone is unperturbed by the local structural adaptations, resulting from the mutation A103L, A103R, or A108L. A direct spectral comparison showed that the chemical shift perturbations that occur on the Skp trimeric state upon t OmpA binding are highly similar for Skp(WT), Skp(A103L), Skp(A103R), and Skp(A108L) (Fig. 5.1.3C, D).

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Next, the authors investigated the effect of tOmpA binding on the Skp monomer-trimer equilibrium at a temperature of 37 °C., where Skp(A108L) is more than 80% in its monomeric state, and Skp(A108R) and Skp(V117P) are completely monomeric (Fig. 5.1.3E, F). For Skp(A108L), binding of tOmpA resulted in a substantial shift of the population levels from the monomeric towards the trimeric state; no change was observed for Skp(A108R) and Skp(V117P) (Fig. 5.1.3E, F).

To determine whether the identified Skp activation mechanism is important under physiologically relevant conditions, mutants were engineered in *Salmonella enterica*. Three of the mutations were selected, two mildest ones A103R and A103L, as well as V117P, and a strain was engineered with complete genetic deletion of the *skp* gene (Δ*skp*). The point mutants nor a complete *skp* deletion affects *Salmonella* fitness in rich lysogeny broth (Fig. 5.1.4A). Then the mutants were tested in competitive infections in a mouse typhoid fever model. In competitive infections, mice are infected with a mixture of WT and mutant strains. Plating bacteria retrieved from the spleen of mice yields the fitness of mutants relative to the WT bacteria. The data reveal a significant fitness defect of *Salmonella skp*(A103L) compared to WT and strong fitness defects for mutants *skp*(A103R) and *skp*(V117P), which are comparable to the complete *skp* deletion (Fig. 5.1.4B).

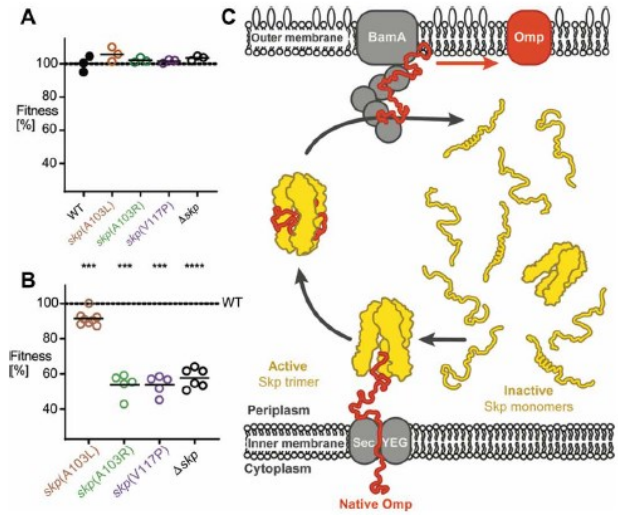

Figure 5.1.4 Essentiality and the functional cycle of the periplasmic chaperone Skp. (A) Fitness of Salmonella strains with various chromosomal skp mutations in rich lysogeny broth. Data for individual cultures and means are shown. (B) Fitness of Salmonella strains in a mouse infection model. Each circle represents data for one mouse from a total of two independent infection experiments (****P < 0.0001 and ***P < 0.001; statistical significance of difference to values for WT based on t test with Holm-Šídák correction for multiple comparisons). (C) Functional cycle of Skp. In the absence of client proteins, Skp populates the periplasm in monomeric form up to low micromolar concentrations. These partially disordered monomers are functionally inactive. An emerging Omp client at the inner membrane recruits an active trimeric chaperone from the ensemble equilibrium. Upon release of the client, trimeric Skp dissociates and the monomers enter the pool of inactive disordered conformations.

These mechanistic insights integrate into an improved picture of the functional cycle of Skp in the bacterial periplasm (Fig. 5.1.4C). Monomeric, disordered Skp molecules populate the periplasmic space. As soon as a client protein emerges from the Sec translocase, the inactive monomers fold and assemble into a trimeric state around the unfolded client protein. Skp directly or indirectly transports the chain to the Bam complex for folding and insertion in the membrane and possibly also to DegP for degradation.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

5.2 Direct observation of a coil-to-helix contraction triggered by vinculin binding to talin

Tapia-Rojo, R., A. Alonso-Caballero, and J. M. Fernandez. 2020. 'Direct observation of a coil-to-helix contraction triggered by vinculin binding to talin', *Sci Adv.* 2020 May 22;6(21):eaaz4707.

Vinculin binds unfolded talin domains in focal adhesion, which recruits actin filaments to reinforce the mechanical coupling of this organelle. The vinculin binding is cooperative, and two vinculin head molecules bind simultaneously to the unfolded R3 domain. To understand how this interaction is regulated, the magnetic tweezers force spectroscopy was used to measure the binding of the vinculin head to the talin R3 domain and study how force modulates this interaction. The vinculin head binding events were resolved as a short contraction of the talin polypeptide due to a coil-to helix transition induced by binding (Fig. 5.2.1).

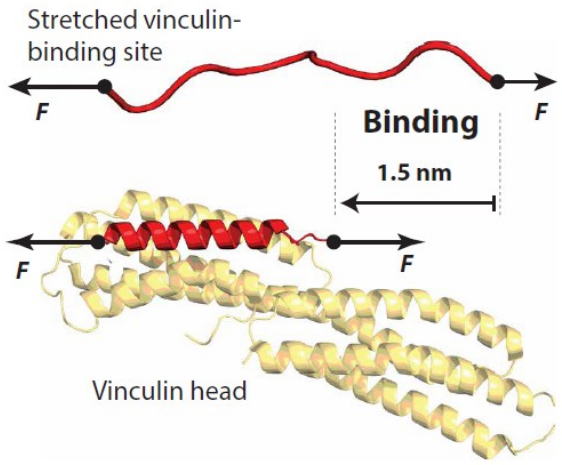

Figure 5.2.1 Vinculin binding requires the structural accommodation of the talin polypeptide on the vinculin head: Under force, talin unfolds and the vinculin-binding sites become unstructured polypeptide chains. Upon binding of one vinculin head molecule, the binding site helix reforms, which shortens talin by ~1.5 nm at a force of 9 pN.

Vinculin binding to the talin R3 domain WT (wild type) and talin R3 IVVI (mutation) was studied. The focus was on the characterization of the mutant because the mutation of four threonines to hydrophobic valine and isoleucine residues increase the domain stability. The custom-made magnetic tweezers were used to apply physiological forces to single R3 domains in the presence of the vinculin head and measure its extension changes in real-time. The molecular construct contains the R3 domain followed by eight repeats of titin I91 domain as molecular handles; a HaloTag enzyme was used for covalent tethering to the glass surface and biotin for anchoring streptavidin-coated superparamagnetic bead (Fig. 5.2.2A). Vinculin head binding to the R3 started from folded R3 IVVI at 4 pN. The increase to 9 pN leads to the exhibition reversible folding/unfolding dynamics of the R3 IVVI

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

domain. Vinculin head binding blocks talin refolding (Fig. 5.2.2B). Using constant force, the binding events were observed as a short contraction of the unfolded polypeptide, which always precedes the arrest of R3 folding dynamics (Fig. 5.2.2B, inset, red arrow). The contraction indicates that vinculin head binding induces a conformational change on unfolded R3, likely a coil-to-helix transition required for the vinculin head to bind its substrate firmly (Fig. 5.2.1). Experiments with R3 WT in the presence of vinculin head reveal this same effect.

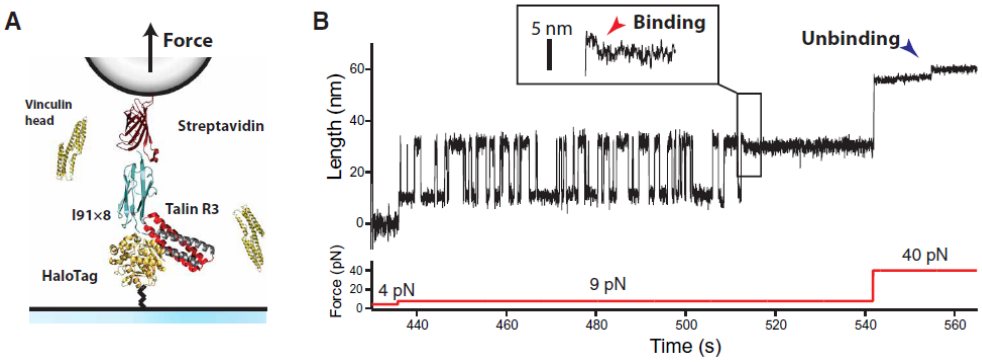

Figure 5.2.2 Real-time detection of the coil-to-helix contraction induced by vinculin head binding. (A) Schematics of a magnetic tweezers experiment for detecting vinculin binding events. The authors engineered an (R3)-(I91)8 protein construct, flanked by a HaloTag for covalent tethering to a glass coverslip, and a biotin for anchoring to streptavidin-coated superparamagnetic beads. Physiological-level forces in the piconewton range are applied through a magnetic field gradient created by either a pair of permanent magnets or a magnetic head. The experiment is conducted in the presence of vinculin head, and the extension changes due to folding or binding are measured with nanometer resolution. (B) Magnetic tweezers recording showing individual vinculin head binding events to the R3 IVVI domain. At 9 pN, R3 IVVI folds and unfolds in equilibrium, which yields extension changes of ~20 nm. In the presence of 20 nM vinculin head, these dynamics eventually stop due to the binding of vinculin head. This event is resolved as a ~3-nm contraction that occurs in the unfolded talin polypeptide due to the reformation of the α-helices of its two vinculin-binding sites (red arrow, inset). The complex dissociates at high forces, showing ~3-nm upward steps (blue arrow).

The mechanical fingerprint for vinculin head binding is a contraction of a few nanometers on the unfolded R3 polypeptide. Vinculin head binding events on R3 IVVI were measured at different forces (Fig. 5.2.3A). The increase of the force leads to an increase in the size of contraction. This is expected in the transition from a random-coiled chain to a compact structure. Although the R3 domain has two vinculin-binding sites, a single contraction event was observed. Two distinct steps with the extension of ~3 nm were resolved when dissociating at forces above 40 pN (Fig. 5.2.3B). Data suggest that two vinculin head molecules bind simultaneously to unfolded R3, whereas each vinculin head unbinds independently. The size of the binding contraction as a function of force was measured (fig. 3C, red circles) and fitted to the FJC (freely jointed chain) model. The obtained contour length change was $\Delta Lc = 7.33 \pm 0.69$ nm (Fig. 5.2.3C). Plotting the same FJC fit with the half contour length results is in agreement with the steps measured for unbinding (Fig. 5.2.3C).

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

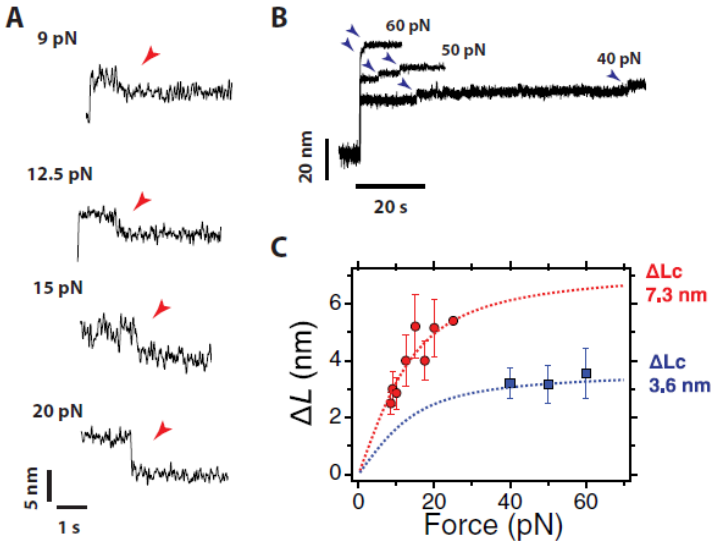

Figure 5.2.3 Vinculin head binding contracts unfolded R3 IVVI. (A) Averaged recordings of the binding contractions at different pulling forces. The magnitude and duration of the contraction depend on the force. Traces averaged from >10 recordings. (B) Unbinding steps at different pulling forces. Two ~3-nm steps are observed, after which talin recovers its ability to refold. (C) Average step sizes for the binding contractions (red) and unbinding steps (blue) measured as a function of the pulling force. The binding contraction scales with force following the FJC polymer model with a contour length of 7.3 nm, which agrees with the simultaneous formation of the two α-helices of the vinculin-binding sites. The steps of unbinding have half that contour length, indicating that they correspond to the unraveling of a single binding site helix. Error bars are the SEM; data collected over 35 molecules, 156 binding steps, and 501 unbinding steps.

For the observation of the talin-bound state, the waiting time was measured (Fig. 5.2.4A). The waiting time t_b is determined from binding trajectories to R3 IVVI at 9 pN and different vinculin head concentrations.

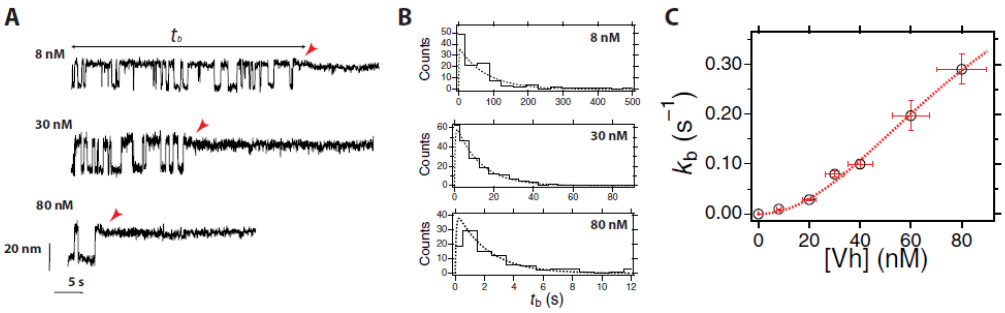

Figure 5.2.4 Stoichiometry of vinculin head binding to R3 IVVI. (A) Magnetic tweezers recordings of R3 IVVI in the presence of 8, 30, and 80 nM vinculin head at a force of 9 pN. The waiting time for vinculin binding (tb) can be measured at the single-molecule level as the time taken since the probe force is set until the contraction is observed and the hopping dynamics stop. (B) Distribution of vinculin head binding times at different concentrations, and a force of 9 pN. The shape of the distribution follows a single-molecule enzymatic model, where two vinculin head molecules bind simultaneously to unfolded R3. (C) Concentration dependence of the rate of vinculin head binding, which follows a second-order Hill-like equation. This demonstrates that two vinculin head molecules bind simultaneously to the unfolded talin R3 domain. The point at 0 nM was estimated from three very long recordings in the absence of vinculin head (5, 7, and 36 hours) that showed no arrest of talin folding dynamics. Vertical error bars are the SEM, and horizontal error bars are the precision on the determination of the vinculin head concentration (12.3%). Number of events measured: 8 nM, N = 122; 20 nM, N = 184; 30 nM, N = 200; 40 nM, N = 267; 60 nM, N = 139; 80 nM, N = 104.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

The distributions of waiting time were fitted (Fig. 5.2.4B). Vinculin head binding kinetics are governed by two competing time scales - the folding/unfolding dynamics of R3 and the concentration-dependent on-rate k^0_{on}. At low concentrations, the on-rate is much slower than R3 folding kinetics, a rate-limiting process by vinculin head association. At higher concentrations, both processes become comparable, but binding cannot occur faster than the R3 domain unfolds. From the distributions was calculated the binding rate k_b, which has a quadratic dependence (Fig. 5.2.4C). Further, vinculin head binding to R3 IVVI was measured at different forces and over a fixed time of 50 s, demonstrating the biphasic effect of force on binding (Fig. 5.2.5A). At 8 pN, R3 IVVI explores the unfolded state with low probability, and vinculin head binds with slow kinetics. As force increases to 9-10 pN, the rate of R3 unfolding increases, leading to faster vinculin head binding. At 15 and 20 pN, R3 IVVI is always unfolded; however, the binding kinetics rapidly slow down, and at 30 pN, binding is blocked. This negative effect of force on binding arises from the coil-to-helix contraction triggered by the binding.

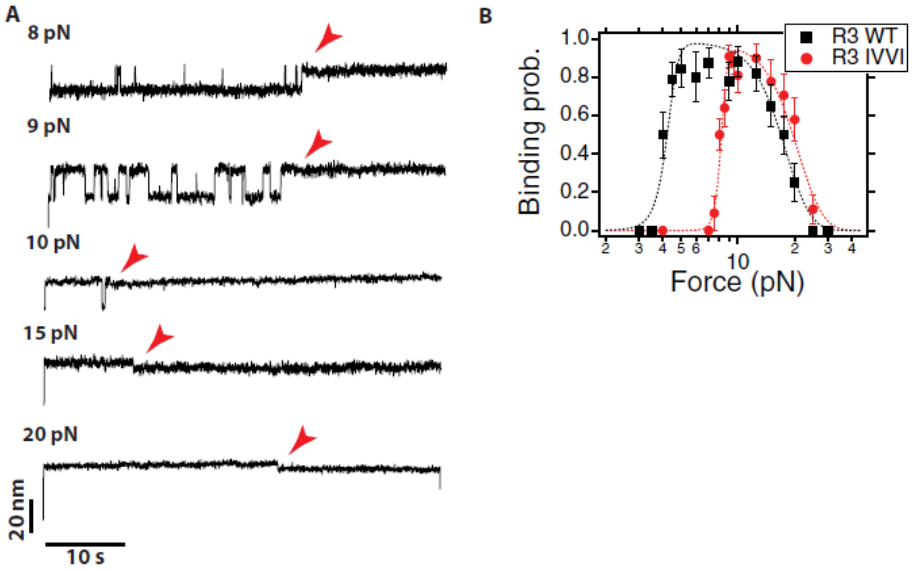

Figure 5.2.5 Mechanical force regulates vinculin head binding. (A) Typical recordings of vinculin head binding to R3 IVVI at different forces during a 50-s time window at a concentration of 20 nM. (B) Binding probability measured over a 50-s time window as a function of force. Force has a biphasic effect on binding, favoring it by unfolding talin, but hampering it due to the energy penalty of the coil-to-helix contraction. The data are described by a simple model based on this mechanism (dashed lines; see section X in the Supplementary Materials). Errors are SEM; data collected over 20 molecules and 259 observations for R3 IVVI, and 10 molecules and 213 observations for R3 WT.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

5.3 Insertion and folding pathways of single membrane proteins guided by translocases and insertases

Serdiuk, T., A. Steudle, S. A. Mari, S. Manioglu, H. R. Kaback, A. Kuhn, and D. J. Muller. 2019.'Insertion and folding pathways of single membrane proteins guided by translocases and insertases', *Sci Adv*,5: eaau6824.

The biogenesis of membrane proteins begins in ribosomes. Next, after exiting a membrane protein from the ribosome, it is necessary to transport proteins into cellular membranes. This step is performed by essential insertases and translocases, which interact with the newly synthesized nascent polypeptide chain and lower the energy barrier for inserting and folding.

This work studied the insertion and folding process of single lactose permease (LacY) precursors assisted by YidC and SecYEG. In bacteria, SecYEG folds α-helical membrane proteins into the inner membrane and translocates precursors of periplasmic and β-barrel outer membrane proteins to the periplasm. YidC inserts membrane proteins into the bacterial inner membrane. SMFS assay was used to pick up the polytopic α-helical LacY from *E. coli* by C-terminal end and to unfold and extract the protein from the membrane mechanically (Fig. 5.3.1A). α-helical membrane proteins exit the ribosome and start insertion and folding from the N-terminal end. For this reason, it was necessary to elongate the C-terminal end of WT LacY with polyGly polypeptide, followed by an eight-amino acid-long His-tag.

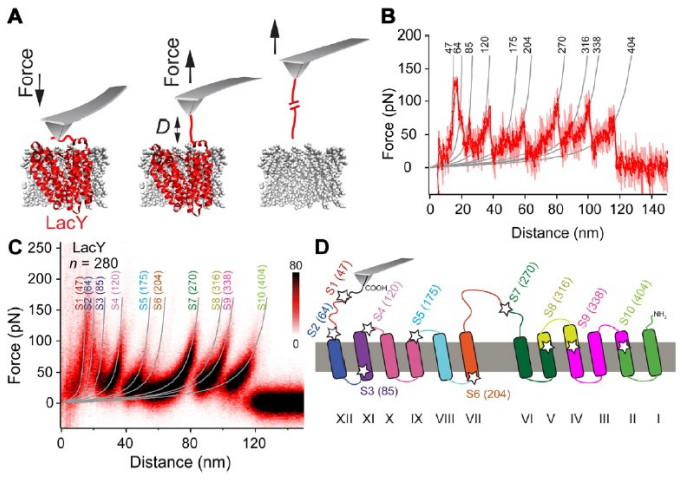

Figure 5.3.1 Unfolding fingerprint pattern of native LacY. (A) Schematics of the mechanical unfolding of native LacY from the phospholipid (PE/PG) membrane. The AFM stylus is pushed onto LacY (PDB 1PV7) to nonspecifically attach the elongated C-terminal end (polyGly LacY). Then, the cantilever is retracted to apply mechanical pulling force to the terminus. During retraction, LacY stepwise unfolds structural segments until being completely unfolded and extracted from the membrane (17). (B) Force-distance curve recorded upon unfolding a single LacY. The force-distance curve is shown as raw data (pale red) and smoothed (Savitzky-Golay filter, dark red). To obtain the contour lengths of mechanically unfolded polypeptide stretches (in amino acid), Authors fit every force peak using the worm-like chain (WLC) model (gray curves; Materials and Methods). (C) Density plot of 280 superimposed force-distance curves each showing the mechanical unfolding of one LacY. Mean contour lengths given at the top of each WLC curve define the ending of the previously unfolded structural segment and the beginning of the next segment to be unfolded. (D) Structural segments S1 to S10 mapped to the secondary structure of LacY as unfolded beginning from the C terminus. The C terminus is shown at the left, and transmembrane α-helices are numbered I to XII.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

To imitate the lipid composition of the *E. coli* inner membrane, the LacY construct was reconstituted into phospholipid membranes. YidC, SecYEG, or YidC and SecYEG were reconstituted into phospholipid liposomes. To ensure that reconstituted YidC and SecYEG were close to each other, a 14-amino acid-long linker was fused on the C-terminal end of SecY and the N-terminal end of YidC. Firtsly, the unfolding fingerprint pattern of native LacY was studied. LacY was attached on the C-terminal end to the AFM stylus (Fig. 5.3.1A), and a force-distance curve was recorded (Fig. 5.3.1A, B). Repeating the experiment led to a highly reproducible force peak pattern of 10 force peaks (Fig. 5.3.1C). Each of the ten force peaks reflects the unfolding of one distinct structural segment of LacY (Fig. 5.3.1D). Force peak pattern is sensitive to the native fold of LacY and changes upon misfolding.

Next, YidC assists LacY polypeptide in inserting and folding into the membrane. The C-terminal end attached to LacY was brought close to the surface of a YidC membrane (Fig. 5.3.2A). A force-distance curve was recorded after a given folding time (range 1-10 s) (Fig. 5.3.2B). The position of each force peak correlates with one out of the ten force peaks in the fingerprint pattern (Fig. 5.3.1C). YidC probably enables the LacY to start the insertion with any of the ten structural segments S1 - S10 (Fig. 5.3.2C).

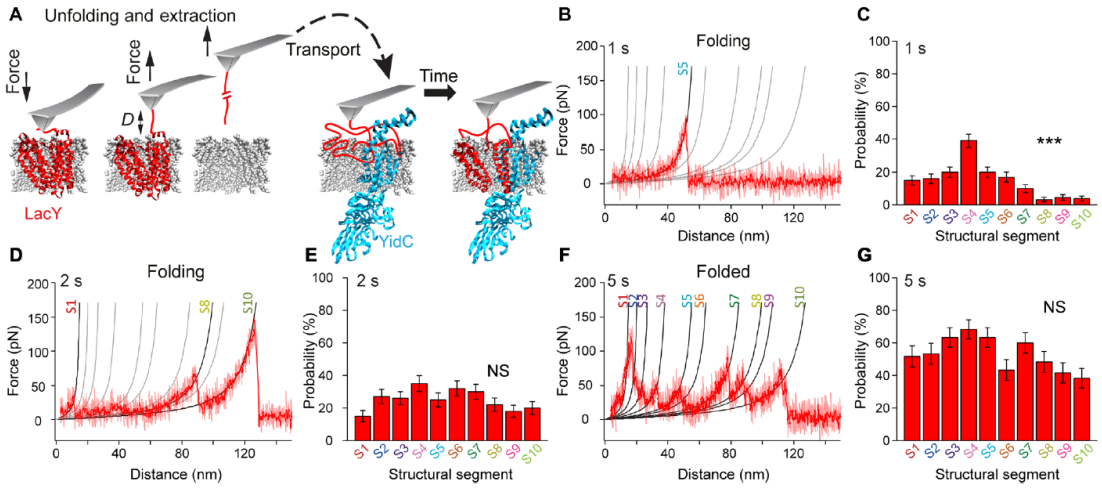

Figure 5.3.2 YidC promotes stochastic insertion of LacY until having folded the native structure. (A) Schematic folding experiment of LacY in the presence of YidC. First, the AFM stylus attached to the C terminus is used to mechanically unfold and extract LacY from the membrane. Then, the unfolded LacY polypeptide is transported by the stylus in close proximity (~5 to 10 nm) to a phospholipid membrane embedding YidC. (B) After a folding time of 1 s, the stylus is retracted, recording a force-distance curve. To reveal whether the polypeptide folded structural segments, force peaks of the curve are fitted with the WLC model. WLC curves matching the fingerprint pattern of native LacY in terms of means ± SD are represented in black, and fits not matching are gray. (C) Probability of structural segments S1 to S10 inserting after 1-s folding time [number of structural segments (nss) = 231]. (D) Force-distance curve recorded in the presence of YidC after 2 s. (E) Probability of segments inserted after 2 s (nss = 266). (F) Force-distance curve recorded in the presence of YidC after 5 s. (G) Probability of segments inserted after 5 s (nss = 319). $\chi 2$ tests indicate (non)uniform distributions (***$P < 0.001$). NS, nonsignificant. Error bars indicate SE.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

At folding time 2 s more force peaks were recorded. Compared with the fingerprint, the polypeptide inserted more structural segments into the YidC membrane (Fig. 5.3.2D). The peaks appear in random order without any preference (Fig. 5.3.2E). At a folding time of 5 s, the number of force peaks (inserted structural segments) increases (Fig. 5.3.2F). This assay was also used for investigating how SecYEG inserts and folds LacY into the membrane. The C-terminus of LacY was mechanically pulling and brought close to a SecYEG membrane (Fig. 5.3.3A). After a folding time of 1 s, the stylus was retracted, and in a force-distance curve (Fig. 5.3.3B) only one force peak is recorded. Repeating experiments led to the same result; however, the single force peak occurred at varying positions. This indicates that the polypeptide could start insertion from any among the ten structural segments of LacY (Fig. 5.3.3C). At 2 s folding time, the LacY inserted about 1 to 3 structural segments (Fig. 5.3.3D). The force peaks were grouped in individual curves, indicating that the polypeptide inserts neighbored segments. The highest probability to insert has S2 and S5 (Fig. 5.3.3E). Folding time 5 s leads to increasing the number of peaks, and at the folding time 10 s, the force-distance curve shows the full fingerprint pattern described for native LacY (Fig. 5.3.3F).

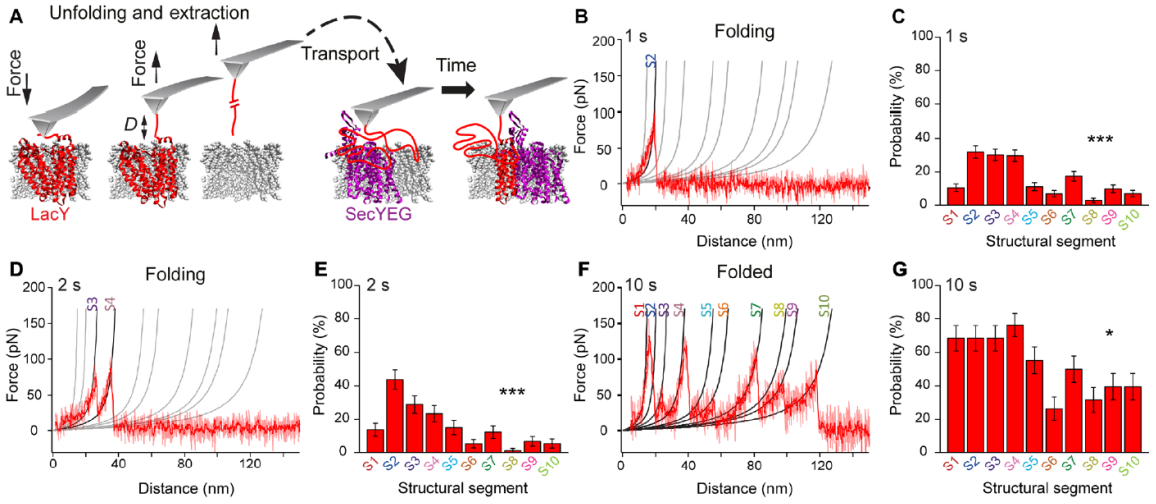

Figure 5.3.3 SecYEG promotes sequential insertion of LacY until having folded the native structure. (A) Schematic LacY folding experiment in the presence of SecYEG. An AFM stylus nonspecifically attached to the C terminus mechanically unfolds and extracts LacY from the membrane. The unfolded LacY polypeptide is then transported by the AFM stylus in close proximity (~5 to 10 nm) to a phospholipid membrane embedding SecYEG where it is kept for 1 s. (B) After this folding time, the AFM stylus is retracted, recording a force-distance curve. To reveal whether the polypeptide folded structural segments, force peaks of the curve are fitted with the WLC model. WLC curves matching the fingerprint pattern of native LacY in terms of means ± SD are represented in black, and fits not matching are gray. The example shows the insertion of structural segment S2. (C) Probability distribution of structural segments inserted after 1 s (nss = 271). (D) Force-distance curve recorded in the presence of SecYEG after 2 s. (E) Probability distribution of segments inserted after 2 s (nss = 114). (F) Force-distance curve recorded in the presence of SecYEG after 10 s. (G) Probability distribution of segments inserted after 10 s (nss = 198). $\chi 2$ tests indicate (non)uniform distributions (*P < 0.05, ***P < 0.001). Error bars indicate SE.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

The effect of SecYEG and YidC together on insertion and folding of the LacY was tested. LacY was brought close to a membrane with SecYEG-YidC construct (Fig. 5.3.4A). At folding time 1 s, force curves showed one force peak (Fig. 5.3.4B). The force peaks varied in position, and the highest priority to insert had S2 and S4 (Fig. 5.3.4C). The increasing folding time to 2 s leads to the detection of more force peaks appearing in groups (Fig. 5.3.4D). The grouped force peaks correlate in position to the unfolding force peaks of native LacY. At folding time 10 s, the polypeptide inserts all ten segments (Fig. 5.3.4F).

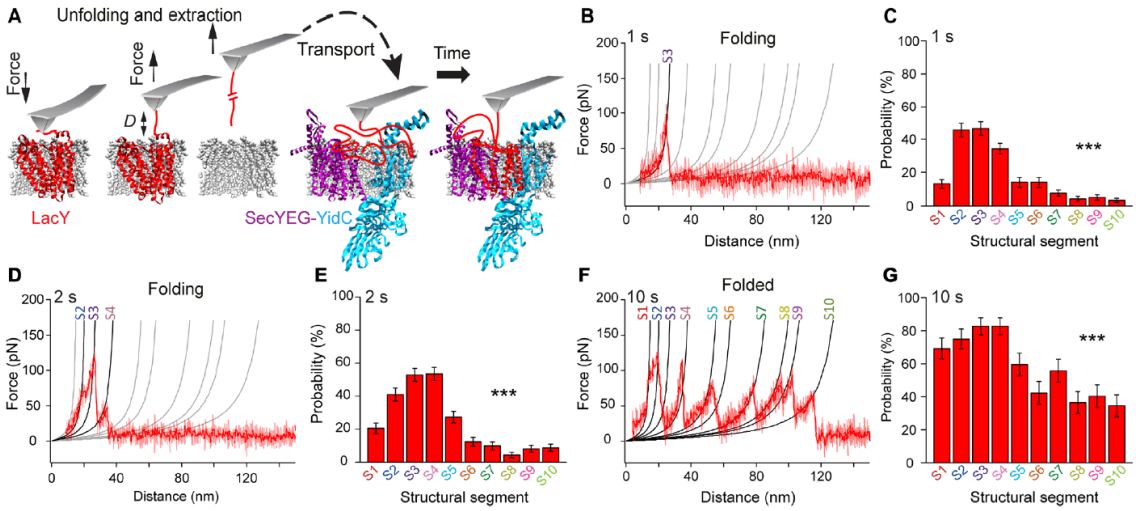

Figure 5.3.4 SecYEG-YidC fusion construct promotes sequential insertion of LacY until having folded the native structure. (A) Schematic LacY folding experiment in the presence of the SecYEG-YidC fusion construct. An AFM stylus nonspecifically attached to the C terminus mechanically unfolds and extracts LacY from the membrane. The unfolded LacY polypeptide is then transported by the AFM stylus in close proximity (~5 to 10 nm) to a phospholipid membrane embedding the SecYEG-YidC construct where it is kept for 1 s. (B) After this folding time, the AFM stylus is retracted, recording a force-distance curve. To reveal whether the polypeptide folded structural segments, force peaks of the curve are fitted with the WLC model. WLC curves matching the fingerprint pattern of native LacY in terms of means ± SD are represented in black, and fits not matching are gray. The example shows the insertion of structural segment S3. (C) Probability distribution of structural segments inserted after 1 s (nss = 206). (D) Force-distance curve recorded in the presence of SecYEG-YidC after 2 s. (E) Probability distribution of segments inserted after 2 s (nss = 384). (F) Force-distance curve recorded in the presence of SecYEG-YidC after 10 s. (G) Probability distribution of segments inserted after 10 s (nss = 301). $\chi 2$ tests indicate (non)uniform distributions (***$P < 0.001$). Error bars indicate SE.

Next, the number of structural segments inserted over time was analyzed (Fig. 5.3.5A). By linear regression, two insertion rates of LacY in the presence of YidC were determined. For folding time ≤ 5 s was 0.96 ± 0.03 segment s^{-1}, and for folding time from 5 to 10 s, it was 0.13 ± 0.08 segment s^{-1}. The similar rate constants 0.42 ± 0.06 segment s^{-1} were determined in the presence of SecYEG and SecYEG-YidC construct for all folding times. This indicates that SecYEG dominates structural segment insertion and folding kinetics over YidC, so YidC is not a limiting factor. Neighbored force peaks indicate the sequential insertion of structural segments (Fig. 5.3.5B). Based on statistical analysis (Fig. 5.3.5C),

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

SecYEG assists the sequential insertion of the LacY polypeptide, and YidC supports the insertion of structural segments in random order.

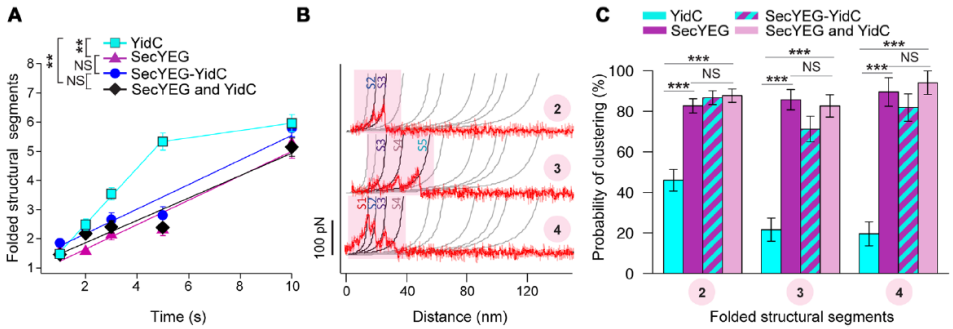

Figure 5.3.5 In the presence of YidC and SecYEG, the translocase defines the folding pathway of the LacY polypeptide. (A) Folding kinetics of LacY in the presence of YidC, SecYEG, SecYEG-YidC fusion construct, or SecYEG and YidC from the SecYEG-YidC construct cleaved by the PreScission protease. Colored linear fits approach the insertion and folding rate of structural segments. (B) Force-distance curves exemplify single LacY polypeptides inserting and folding 2, 3, or 4 neighbored structural segments in the presence of SecYEG. Force peaks matching the fingerprint pattern of native LacY are represented in black, and fits not matching are gray. (C) Probabilities of detecting the insertion of 2, 3, or 4 neighbored segments in the presence of YidC, SecYEG, SecYEG-YidC, or SecYEG and YidC. Four hundred seventy-eight experiments (force-distance curves) detecting insertion and folding events of LacY have been recorded in the presence of YidC, 395 in the presence of SecYEG, 397 in the presence of SecYEG-YidC fusion construct, and 313 in the presence of SecYEG and YidC. Statistical differences examined by analysis of covariance (A) and Z (C) tests were considered nonsignificant for $P > 0.05$ and significant for $^{**}P < 0.01$ and $^{***}P < 0.001$. Error bars indicate SE.

Independently of the insertion and folding mechanism, the translocon and insertase could complete the folding process independently or cooperatively. Many membrane proteins can use YidC or SecYEG for insertion and folding, but insertase inserted structural segments in random order, the translocon inserted them sequentially. The random insertion of segments allows 10! (3 628 800) possible pathways, while the sequential insertion leads to only ten pathways (Fig. 5.3.6).

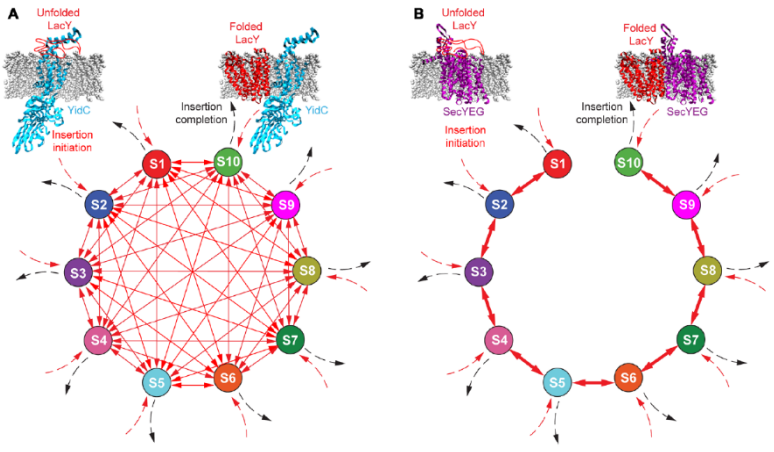

Figure 5.3.6 SecYEG and YidC insert and fold the membrane protein LacY along different pathways. (A) YidC starts insertion of the LacY polypeptide from any structural segment after which it stepwise inserts the remaining structural segments S1 to S10 of LacY. Structural segments insert in random order until folding of LacY has been completed. YidC offers 10! (3,628,800) pathways to fold the 10 structural segments S1 to S10 of LacY toward the native structure. (B) SecYEG alone or SecYEG and YidC together start insertion of the LacY polypeptide from any structural segment after which the remaining segments are inserted sequentially until folding of LacY has been completed. SecYEG or SecYEG and YidC offer 10 principal pathways to fold the 10 structural segments toward native LacY. Red dashed arrows indicate possibilities of initiating insertion. Red double arrows highlight the insertion and folding steps of the folding pathways. Black dashed arrows indicate the completion of the insertion and folding process.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

5.4 A network of chaperones prevents and detects failures in membrane protein-lipid bilayer integration

Coelho, J. P. L., M. Stahl, N. Bloemeke, K. Meighen-Berger, C. P. Alvira, Z. R. Zhang, S. A. Sieber, and M. J. Feige. 2019. 'A network of chaperones prevents and detects failures in membrane protein-lipid bilayer integration', *Nat Commun*, 10: 672.

A critical step in the biosynthesis of membrane proteins is their correct topogenesis, integrating each TM helix in the adequate orientation into the lipid bilayer. It remains unclear how cells detect and handle failures in this process.

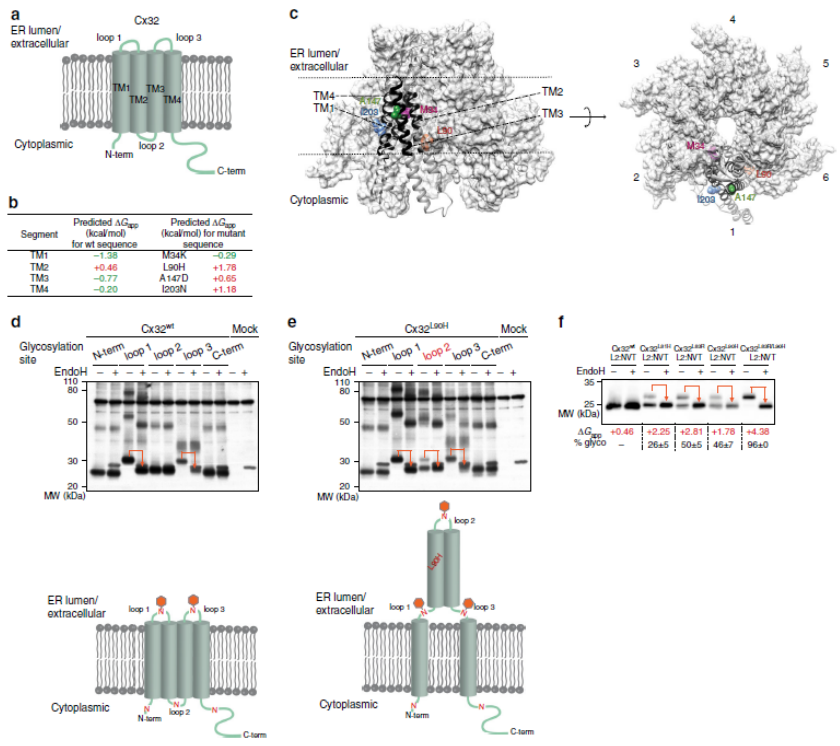

Figure 5.4.1 Single-point mutations lead to failures in membrane integration for Cx32. a Schematic of Cx32, showing its predicted topology. b Predicted free energies for helix insertion for apolar-to-polar missense mutations investigated in this study. For $\Delta G < 0$ (green) an energetically favorable membrane integration is predicted, while $\Delta G > 0$ (red) indicates an unfavorable insertion reaction. Transmembrane helices and ΔG values were predicted according to Hessa *et al.*17. c Side and top view of the modeled hexameric Cx32 connexon. Disease-causing mutants investigated in this study are shown in a CPK representation on a single monomer. Transmembrane helices are shown in black. Individual Cx32 monomers are numbered from 1 to 6. d Cx32wt and e Cx32L90H with individually introduced glycosylation sites in the indicated regions were transfected into HEK293T cells, lysates treated with or without EndoH as indicated and analyzed by immunoblotting. The schematics below each indicate the location of the individually assessed glycosylation sites (N) and possible topologies deduced from the observed glycosylation. An orange arrow in the blots and an orange hexagon in the schematic indicate sites that became glycosylated. f HEK293T cells transfected with the indicated Cx32 TM2 mutant constructs were analyzed as in d, e. Constructs all carried a glycosylation reporter site in loop 2 (L2:NVT). Predicted free energies of membrane insertion for TM2 of the respective proteins as well as a quantification of relative loop 2 reporter site glycosylation (mean ± SEM, N = 3) are shown below the immunoblots

Mutations in the membrane proteins are connected with many diseases; an example is a disorder of the peripheral nervous system Charcot-Marie-Tooth disease, which is caused by single point mutations in the membrane protein Cx32. The gap junctions forming protein Cx32 contains four TM helices, with its N- and C-termini exposed to the cytoplasm (Fig. 5.4.1a). To clarify how mutations

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

affect the topogenesis, four disease-causing mutations within Cx32 were selected, one in each TM helices, all predicted to significantly destabilize integration of the respective helix (Fig. 5.4.1b). Based on the crystal structure of connexin 26, a model of Cx32 was generated for defining the exact location of each mutation; all of the four mutations were predicted to be located within the membrane (Fig. 5.4.1c).

Next, an artificial glycosylation site was included in each of the three Cx32 loops connecting its four TM helices as well as its C-terminal tail (Fig. 5.4.1d, e). Into the N-terminal tail of Cx32 was omitted artificial glycosylation. In mutant Cx32^{L90H}, partial glycosylation of the reporter site in loop 2 (Fig. 5.4.1e) was detected, arguing for its exposure to the ER lumen. It suggests that Cx32^{L90H} can adopt an altered topology, with TM helices 2 and 3 becomings exposed to the ER lumen. To asses, if these findings are general or limited to Cx32^{L90H}, two additional disease-associated mutants of Cx32 affecting TM helix 2 were analyzed. Both of these mutants also showed glycosylation of a loop 2 reporter site. It suggests that mutation in the TM helix 2 leads to membrane disintegration and failures in topogenesis (Fig. 5.4.1f).

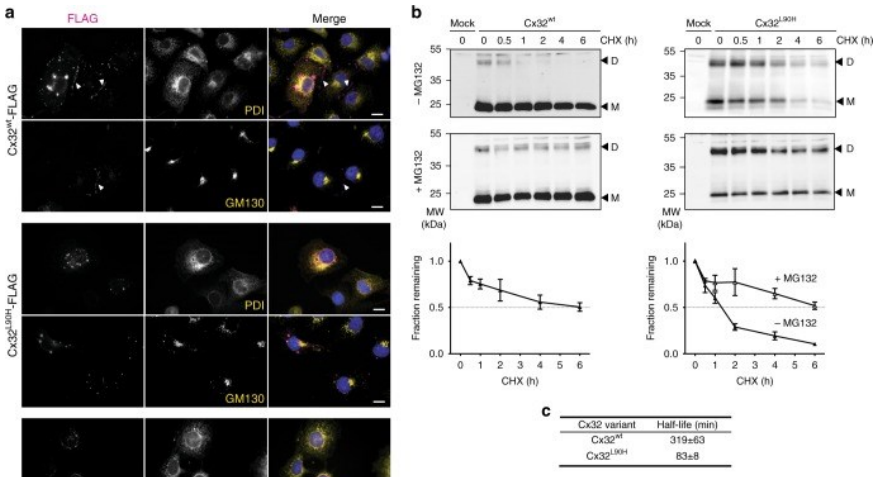

Figure 5.4.2 Cx32L90H shows defects in gap-junction formation and rapid degradation. a COS-7 cells were transfected with the indicated constructs and immunostained for FLAG-tagged Cx32 (magenta), PDI (yellow) as an ER marker, or GM130 (yellow) as a Golgi marker. Nuclei were stained with DAPI (blue). Anti-FLAG immunofluorescence data are depicted as maximum intensity projections from deconvoluted z-stacks, while PDI, GM130, and nuclei are shown as a central cell plane from the same, nondeconvoluted images. Gap-junction plates lining cell–cell boundaries are indicated with white arrowheads. Pictures are representative of cells from at least three different biological replicates. Scale bars correspond to 20 µm. b HEK293T cells transfected with the indicated constructs were incubated with either cycloheximide (CHX) alone, or additionally with the proteasome inhibitor MG132 where indicated. Arrowheads indicate monomer (M) and dimer (D) bands quantified to determine Cx32 turnover. Quantifications are shown below the immunoblots (mean ± SEM, N ≥ 3). c Half-lifes without MG132 for Cx32wt and Cx32L90H as derived from b are shown

Since the failures in membrane integration for Cx32^{L92H} were observed, the recognition of the mutants by ER quality control was next studied. Microscopy experiments for Cx32^{wt} revealed fluorescent punctae in transfected cells, frequently lining cell-cell boundaries (Fig. 5.4.2a). Mutant

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Cx32^{L90H} formed some punctae; however, these were reduced in number in comparison to Cx32^{wt} and appeared not to be localized at the typical cell-cell boundary observed for Cx32^{wt} (Fig. 5.4.2a). Cx32^{L90H} did not form detergent-resistant species, which were kept for Cx32^{wt} and characteristic of gap-junction plaques.

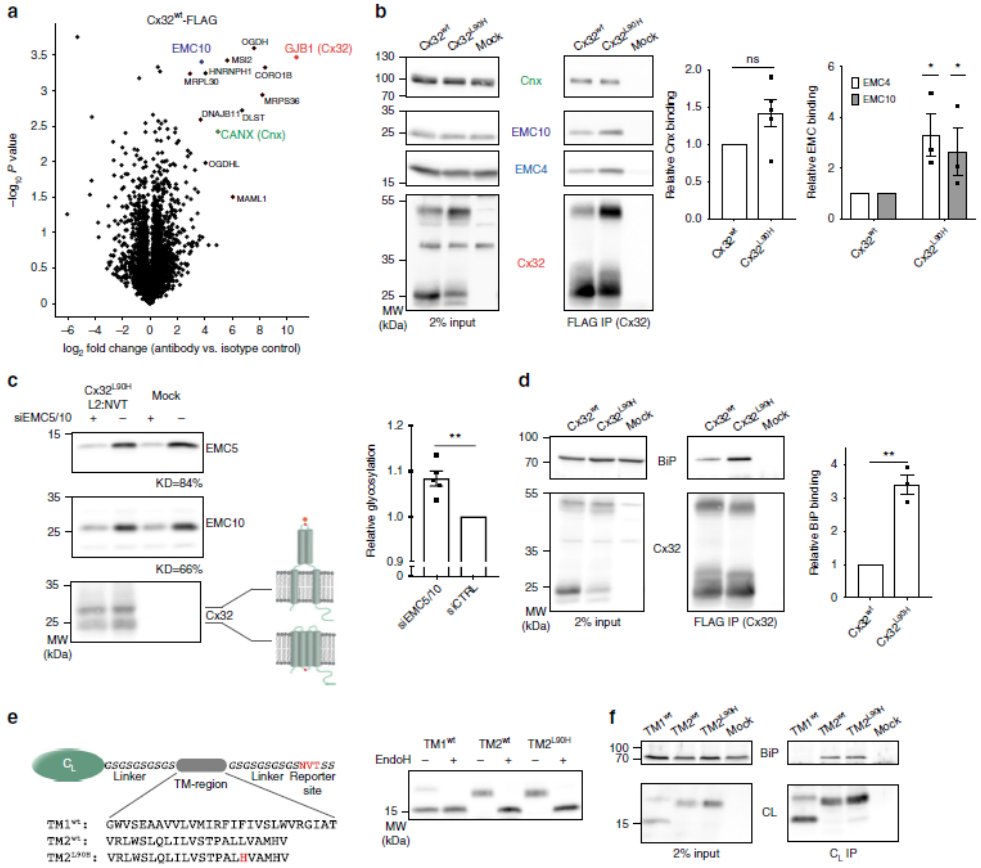

Figure 5.4.3 Connexin mutants are recognized by the ER quality control system. a Mass spectrometry volcano plot for FLAG-tagged Cx32wt, immunoprecipitated in 1% digitonin from transfected HEK293T cells. Enriched proteins are denoted with their Uniprot gene name. Cx32 is shown in orange and ER chaperones investigated further in this study are highlighted in blue (EMC10) and green (Cnx). Either a rabbit monoclonal anti-FLAG antibody or a rabbit IgG isotype control was used. b Representative blots from immunoprecipitation experiments from HEK293T cells transfected with the indicated Cx32 constructs. Interaction of Cx32 with endogenous Cnx and EMC subunits was detected and increased for Cx32L90H with both EMC4 and EMC10 (mean ± SEM, N ≥ 3, ns: nonsignificant, *P value < 0.05, two-tailed Student's t tests). Quantifications were performed as described in the Methods section. c Transient knockdown of EMC5/10 by siRNA (average knockdown (KD) efficiencies are shown below the blots) increases glycosylation of a reporter site in loop 2 for Cx32L90H. Monomeric species ±glycosylation are shown on the blot, indicative of the topologies depicted on the right. Changes in glycosylation, quantified as described in the Methods section, are shown on the right (mean ± SEM, N = 5, **P value < 0.01, two-tailed Student's t tests). d Same as in b for co-transfected hamster BiP (mean ± SEM, N = 3, **P value < 0.01, two-tailed Student's t tests). e Schematic of a reporter construct to assess BiP-TM region binding. The nonglycosylated immunoglobulin λ light chain CL domain with its own ER import sequence is followed by a flexible linker connected to the TM sequence of interest shown below the schematic. TM segment 1 was inverted to allow for a type I topology. A C-terminal glycosylation site (NVT, marked in red) allows to assess membrane integration (no glycosylation) versus ER import (possible glycosylation). Membrane integration/ER import was assessed for Cx32 TM segment 1, 2, and 2 carrying the L90H mutation by transfection of the constructs into HEK293T cells and EndoH deglycosylation where indicated. f CL-TM constructs were co-transfected with hamster BiP into HEK293T cells and their interaction was analyzed by co-immunoprecipitation experiments coupled to immunoblots.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Next, it was investigated if quality control processes would recognize Cx32L90H and become a substrate of ERAD. $Cx32^{wt}$ was degraded with a half-life of 5-6 h, and mutant $Cx32^{L90H}$ ca. 1.5 h (Fig. 5.4.2b,c), and the proteasome inhibitor carbobenzoxy-Leu-Leu-leucinal (MG132) blocked the degradation of mutant. Further, how the mutation affected chaperone interactions in the ER was detected. FLAG-tagged $Cx32^{wt}$ and $Cx32^{L90H}$ were transfected into HEK293T cells, and cells were lysed in mild digitonin buffer; cx32 proteins were immunoprecipitated performed affinity-enrichment MS. No ER chaperone interaction was either exclusively found for the wild type or the mutant; the MS revealed two potential interaction partners of Cx32 - the ER chaperone Calnexin (Cnx) and the ER-membrane protein complex subunit 10 (EMC10) (Fig. 5.4.3a). Co-immunoprecipitation experiments revealed interaction for $Cx32^{wt}$ and $Cx32^{L90H}$ with only slightly stronger interaction of the mutant with Cnx (Fig. 5.4.3b). Endogenous EMC4 and EMC10 showed significantly stronger interaction with $Cx32^{L90H}$ than $Cx32^{wt}$ (Fig. 5.4.3b). Increased EMC interaction was also observed for another mutant, which showed membrane misintegration ($Cx32^{L81H}$).

EMC participates in the biogenesis of tail-anchored and multipass TM proteins; for understanding biochemical effects, knockdown of EMC5 and 10on membrane integration of Cx32 was researched. Knockdown by ca. 65-85% led to a small but significant increase in glycosylation of the $Cx32^{L90H}$ mutant at the site in loop 2 (Fig. 5.4.3c). The same behavior was observed for $Cx32^{L81H}$, showing that this effect of EMC is more general. Failures in membrane integration can expose TM helices of Cx32 to the ER lumen. Therefore next step was studying if ER-lumenal chaperone BiP would be able to recognize misintegrated TM helices of $Cx32^{L90H}$.

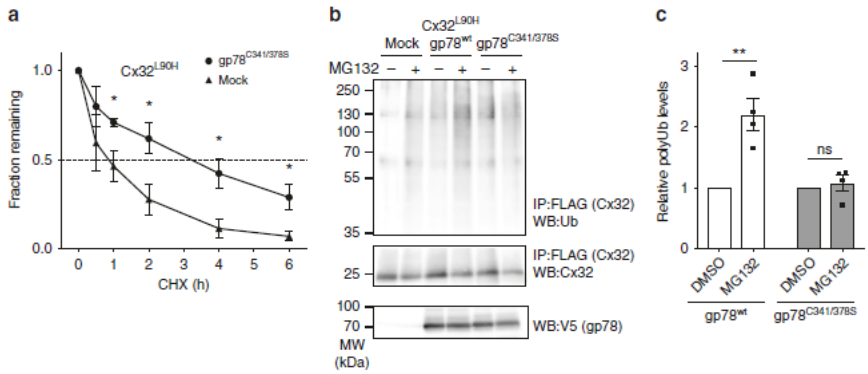

Figure 5.4.4 gp78 mediates Cx32 degradation. a Overexpression of a nonfunctional gp78 mutant (C341/378S) significantly decelerates Cx32L90H degradation (mean ± SEM, N ≥ 3, *P value < 0.05, two-tailed Student's t tests). b Mutant gp78 (C341/378S) overexpression inhibits polyubiquitination of Cx32. FLAG-tagged Cx32L90H was expressed either alone or in the presence of V5-tagged wild type gp78 or gp78C341/378S as indicated. Samples were immunoprecipitated against FLAG (Cx32L90H) and blotted against Cx32 or ubiquitin (Ub). Where shown, MG132 was added to inhibit proteasomal degradation of Ub-modified Cx32L90H. As can be seen from the blots and quantifications in c, MG132 leads to increased ubiquitination of Cx32L90H in the absence or presence of gp78wt overexpression, while gp78C341/378S overexpression inhibits Cx32L90H polyubiquitination. c Quantifications of the data shown in b (mean ± SEM, N = 4, ns: nonsignificant, **P value < 0.01, two-tailed Student's t tests).

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

BiP bound significantly stronger to Cx32^{L90H} and Cx32^{L81H} than to Cx32^{wt} (Fig. 5.4.3d). TM segments of interest were fused to the BiP-inert antibody C_L domain (Fig. 5.4.3e). Only a very weak interaction between TM segment 1 and BiP was observed, TM segment 2 (wild type and mutant) strongly bound to BiP, corroborating that these TM sequences can be directly recognized by the ER-lumenal chaperone BiP (Fig. 5.4.3f). Next was investigated which ERAD are involved in degrading Cx32 mutants. Overexpression of the inactive gp78 RING finger mutant C341/378S has a much stronger stabilizing effect on Cx32^{L90H} (Fig. 5.4.4a). Overexpression of this reduced mutant Cx32^{L90H}ubiquitination upon inhibition of the proteasome with MG132 (Fig. 5.4.4b,c). It suggests that gp78 plays a significant role in the ERAD of Cx32^{L90H} as a membrane protein with an aberrant topology.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

6. REFERENCES

Abdel-Nour, M., L. A. M. Carneiro, J. Downey, J. Tsalikis, A. Outlioua, D. Prescott, L. S. Da Costa, E. S. Hovingh, A. Farahvash, R. G. Gaudet, R. Molinaro, R. van Dalen, C. C. Y. Lau, F. C. Azimi, N. K. Escalante, A. Trotman-Grant, J. E. Lee, S. D. Gray-Owen, M. Divangahi, J. J. Chen, D. J. Philpott, D. Arnoult, and S. E. Girardin. 2019. 'The heme-regulated inhibitor is a cytosolic sensor of protein misfolding that controls innate immune signaling', *Science*, 365.

Alevra, M., S. Mandad, T. Ischebeck, H. Urlaub, S. O. Rizzoli, and E. F. Fornasiero. 2019. 'A mass spectrometry workflow for measuring protein turnover rates in vivo', *Nat Protoc*, 14: 3333-65.

Alley, E. C., G. Khimulya, S. Biswas, M. AlQuraishi, and G. M. Church. 2019. 'Unified rational protein engineering with sequence-based deep representation learning', *Nat Methods*, 16: 1315-22.

Alonso-Caballero, A., D. J. Echelman, R. Tapia-Rojo, S. Haldar, E. C. Eckels, and J. M. Fernandez. 2021. 'Protein folding modulates the chemical reactivity of a Gram-positive adhesin', *Nat Chem*, 13: 172-81.

Altenburg, W. J., N. A. Yewdall, D. F. M. Vervoort, Mhme van Stevendaal, A. F. Mason, and J. C. M. van Hest. 2020. 'Programmed spatial organization of biomacromolecules into discrete, coacervate-based protocells', *Nat Commun*, 11: 6282.

Aprile-Garcia, F., P. Tomar, B. Hummel, A. Khavaran, and R. Sawarkar. 2019. 'Nascent-protein ubiquitination is required for heat shock-induced gene downregulation in human cells', *Nat Struct Mol Biol*, 26: 137-46.

Audet, M., K. Villers, J. Velasquez, M. Chu, C. Hanson, and R. C. Stevens. 2020. 'Small-scale approach for precrystallization screening in GPCR X-ray crystallography', *Nat Protoc*, 15: 144-60.

Behring, J. B., S. van der Post, A. D. Mooradian, M. J. Egan, M. I. Zimmerman, J. L. Clements, G. R. Bowman, and J. M. Held. 2020. 'Spatial and temporal alterations in protein structure by EGF regulate cryptic cysteine oxidation', *Sci Signal*, 13.

Benhaim, M. A., V. Mangala Prasad, N. K. Garcia, M. Guttman, and K. K. Lee. 2020. 'Structural monitoring of a transient intermediate in the hemagglutinin fusion machinery on influenza virions', *Sci Adv*, 6: eaaz8822.

Bhattacharya, K., L. Weidenauer, T. M. Luengo, E. C. Pieters, P. C. Echeverria, L. Bernasconi, D. Wider, Y. Sadian, M. B. Koopman, M. Villemin, C. Bauer, S. G. D. Rudiger, M. Quadroni, and D. Picard. 2020. 'The Hsp70-Hsp90 co-chaperone Hop/Stip1 shifts the proteostatic balance from folding towards degradation', *Nat Commun*, 11: 5975.

Bindels, D. S., M. Postma, L. Haarbosch, L. van Weeren, and T. W. J. Gadella, Jr. 2020. 'Multiparameter screening method for developing optimized red-fluorescent proteins', *Nat Protoc*, 15: 450-78.

Boyken, S. E., M. A. Benhaim, F. Busch, M. Jia, M. J. Bick, H. Choi, J. C. Klima, Z. Chen, C. Walkey, A. Mileant, A. Sahasrabuddhe, K. Y. Wei, E. A. Hodge, S. Byron, A. Quijano-Rubio, B. Sankaran, N. P. King, J. Lippincott-Schwartz, V. H. Wysocki, K. K. Lee, and D. Baker. 2019. 'De novo design of tunable, pH-driven conformational changes', *Science*, 364: 658-64.

Butera, D., and P. J. Hogg. 2020. 'Fibrinogen function achieved through multiple covalent states', *Nat Commun*, 11: 5468.

Campos, L. A., R. Sharma, S. Alvira, F. M. Ruiz, B. Ibarra-Molero, M. Sadqi, C. Alfonso, G. Rivas, J. M. Sanchez-Ruiz, A. Romero Garrido, J. M. Valpuesta, and V. Munoz. 2019. 'Engineering protein assemblies with allosteric control via monomer fold-switching', *Nat Commun*, 10: 5703.

Cao, L., I. Goreshnik, B. Coventry, J. B. Case, L. Miller, L. Kozodoy, R. E. Chen, L. Carter, A. C. Walls, Y. J. Park, E. M. Strauch, L. Stewart, M. S. Diamond, D. Veesler, and D. Baker. 2020. 'De novo design of picomolar SARS-CoV-2 miniprotein inhibitors', *Science*, 370: 426-31.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Cao, Q., D. R. Boyer, M. R. Sawaya, P. Ge, and D. S. Eisenberg. 2019. 'Cryo-EM structures of four polymorphic TDP-43 amyloid cores', *Nat Struct Mol Biol*, 26: 619-27.

Carrasco-Lopez, C., E. M. Zhao, A. A. Gil, N. Alam, J. E. Toettcher, and J. L. Avalos. 2020. 'Development of light-responsive protein binding in the monobody non-immunoglobulin scaffold', *Nat Commun*, 11: 4045.

Carroll, E. C., E. R. Greene, A. Martin, and S. Marqusee. 2020. 'Site-specific ubiquitination affects protein energetics and proteasomal degradation', *Nat Chem Biol*, 16: 866-75.

Chen, K. M., D. Keri, and P. Barth. 2020. 'Computational design of G Protein-Coupled Receptor allosteric signal transductions', *Nat Chem Biol*, 16: 77-86.

Chen, M., X. Chen, N. P. Schafer, C. Clementi, E. A. Komives, D. U. Ferreiro, and P. G. Wolynes. 2020. 'Surveying biomolecular frustration at atomic resolution', *Nat Commun*, 11: 5944.

Chen, X., B. Ji, X. Hao, X. Li, F. Eisele, T. Nystrom, and D. Petranovic. 2020. 'FMN reduces Amyloid-beta toxicity in yeast by regulating redox status and cellular metabolism', *Nat Commun*, 11: 867.

Chen, X., R. Li, S. H. D. Wong, K. Wei, M. Cui, H. Chen, Y. Jiang, B. Yang, P. Zhao, J. Xu, H. Chen, C. Yin, S. Lin, W. Y. Lee, Y. Jing, Z. Li, Z. Yang, J. Xia, G. Chen, G. Li, and L. Bian. 2019. 'Conformational manipulation of scale-up prepared single-chain polymeric nanogels for multiscale regulation of cells', *Nat Commun*, 10: 2705.

Chen, X., S. Liao, Y. Makaros, Q. Guo, Z. Zhu, R. Krizelman, K. Dahan, X. Tu, X. Yao, I. Koren, and C. Xu. 2021. 'Molecular basis for arginine C-terminal degron recognition by Cul2(FEM1) E3 ligase', *Nat Chem Biol*, 17: 254-62.

Chen, X., N. Rajasekaran, K. Liu, and C. M. Kaiser. 2020. 'Synthesis runs counter to directional folding of a nascent protein domain', *Nat Commun*, 11: 5096.

Chen, Z., R. D. Kibler, A. Hunt, F. Busch, J. Pearl, M. Jia, Z. L. VanAernum, B. I. M. Wicky, G. Dods, H. Liao, M. S. Wilken, C. Ciarlo, S. Green, H. El-Samad, J. Stamatoyannopoulos, V. H. Wysocki, M. C. Jewett, S. E. Boyken, and D. Baker. 2020. 'De novo design of protein logic gates', *Science*, 368: 78-84.

Chitwood, P. J., and R. S. Hegde. 2020. 'An intramembrane chaperone complex facilitates membrane protein biogenesis', *Nature*, 584: 630-34.

Choi, H. K., D. Min, H. Kang, M. J. Shon, S. H. Rah, H. C. Kim, H. Jeong, H. J. Choi, J. U. Bowie, and T. Y. Yoon. 2019. 'Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway', *Science*, 366: 1150-56.

Chui, A. J., M. C. Okondo, S. D. Rao, K. Gai, A. R. Griswold, D. C. Johnson, D. P. Ball, C. Y. Taabazuing, E. L. Orth, B. A. Vittimberga, and D. A. Bachovchin. 2019. 'N-terminal degradation activates the NLRP1B inflammasome', *Science*, 364: 82-85.

Coelho, J. P. L., M. Stahl, N. Bloemeke, K. Meighen-Berger, C. P. Alvira, Z. R. Zhang, S. A. Sieber, and M. J. Feige. 2019. 'A network of chaperones prevents and detects failures in membrane protein lipid bilayer integration', *Nat Commun*, 10: 672.

Collier, M. P., T. R. Alderson, C. P. de Villiers, D. Nicholls, H. Y. Gastall, T. M. Allison, M. T. Degiacomi, H. Jiang, G. Mlynek, D. O. Furst, P. F. M. van der Ven, K. Djinovic-Carugo, A. J. Baldwin, H. Watkins, K. Gehmlich, and J. L. P. Benesch. 2019. 'HspB1 phosphorylation regulates its intramolecular dynamics and mechanosensitive molecular chaperone interaction with filamin C', *Sci Adv*, 5: eaav8421.

Cong, Q., I. Anishchenko, S. Ovchinnikov, and D. Baker. 2019. 'Protein interaction networks revealed by proteome coevolution', *Science*, 365: 185-89.

Courtney, T. M., and A. Deiters. 2019. 'Optical control of protein phosphatase function', *Nat Commun*, 10: 4384.

Coyote-Maestas, W., Y. He, C. L. Myers, and D. Schmidt. 2019. 'Domain insertion permissibility-guided engineering of allostery in ion channels', *Nat Commun*, 10: 290.

Dawson, J. E., A. Bah, Z. Zhang, R. M. Vernon, H. Lin, P. A. Chong, M. Vanama, N. Sonenberg, C. C. Gradinaru, and J. D. Forman-Kay. 2020. 'Non-cooperative 4E-BP2 folding with exchange

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

between eIF4E-binding and binding-incompatible states tunes cap-dependent translation inhibition', *Nat Commun*, 11: 3146.

Dijkman, P. M., J. C. Munoz-Garcia, S. R. Lavington, P. S. Kumagai, R. I. Dos Reis, D. Yin, P. J. Stansfeld, A. J. Costa-Filho, and A. Watts. 2020. 'Conformational dynamics of a G protein-coupled receptor helix 8 in lipid membranes', *Sci Adv*, 6: eaav8207.

Ding, X., Z. Zou, and C. L. Brooks Iii. 2019. 'Deciphering protein evolution and fitness landscapes with latent space models', *Nat Commun*, 10: 5644.

Dishman, A. F., R. C. Tyler, J. C. Fox, A. B. Kleist, K. E. Prehoda, M. M. Babu, F. C. Peterson, and B. F. Volkman. 2021. 'Evolution of fold switching in a metamorphic protein', *Science*, 371: 86-90.

Dziekan, J. M., G. Wirjanata, L. Dai, K. D. Go, H. Yu, Y. T. Lim, L. Chen, L. C. Wang, B. Puspita, N. Prabhu, R. M. Sobota, P. Nordlund, and Z. Bozdech. 2020. 'Cellular thermal shift assay for the identification of drug-target interactions in the Plasmodium falciparum proteome', *Nat Protoc*, 15: 1881-921.

Ebo, J. S., J. C. Saunders, P. W. A. Devine, A. M. Gordon, A. S. Warwick, B. Schiffrin, S. E. Chin, E. England, J. D. Button, C. Lloyd, N. J. Bond, A. E. Ashcroft, S. E. Radford, D. C. Lowe, and D. J. Brockwell. 2020. 'An in vivo platform to select and evolve aggregation-resistant proteins', *Nat Commun*, 11: 1816.

Eftekharzadeh, B., V. C. Banduseela, G. Chiesa, P. Martinez-Cristobal, J. N. Rauch, S. R. Nath, D. M. C. Schwarz, H. Shao, M. Marin-Argany, C. Di Sanza, E. Giorgetti, Z. Yu, R. Pierattelli, I. C. Felli, I. Brun-Heath, J. Garcia, A. R. Nebreda, J. E. Gestwicki, A. P. Lieberman, and X. Salvatella. 2019. 'Hsp70 and Hsp40 inhibit an inter-domain interaction necessary for transcriptional activity in the androgen receptor', *Nat Commun*, 10: 3562.

Ellison, C. K., T. N. Dalia, A. B. Dalia, and Y. V. Brun. 2019. 'Real-time microscopy and physical perturbation of bacterial pili using maleimide-conjugated molecules', *Nat Protoc*, 14: 1803-19.

Emond, S., M. Petek, E. J. Kay, B. Heames, S. R. A. Devenish, N. Tokuriki, and F. Hollfelder. 2020. 'Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis', *Nat Commun*, 11: 3469.

Escobedo, A., B. Topal, M. B. A. Kunze, J. Aranda, G. Chiesa, D. Mungianu, G. Bernardo-Seisdedos, B. Eftekharzadeh, M. Gairi, R. Pierattelli, I. C. Felli, T. Diercks, O. Millet, J. Garcia, M. Orozco, R. Crehuet, K. Lindorff-Larsen, and X. Salvatella. 2019. 'Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor', *Nat Commun*, 10: 2034.

Faull, S. V., E. L. K. Elliston, B. Gooptu, A. M. Jagger, I. Aldobiyan, A. Redzej, M. Badaoui, N. Heyer-Chauhan, S. T. Rashid, G. M. Reynolds, D. H. Adams, E. Miranda, E. V. Orlova, J. A. Irving, and D. A. Lomas. 2020. 'The structural basis for Z alpha1-antitrypsin polymerization in the liver', *Sci Adv*, 6.

Faust, O., M. Abayev-Avraham, A. S. Wentink, M. Maurer, N. B. Nillegoda, N. London, B. Bukau, and R. Rosenzweig. 2020. 'HSP40 proteins use class-specific regulation to drive HSP70 functional diversity', *Nature*, 587: 489-94.

Favretto, F., D. Flores, J. D. Baker, T. Strohaker, L. B. Andreas, L. J. Blair, S. Becker, and M. Zweckstetter. 2020. 'Catalysis of proline isomerization and molecular chaperone activity in a tug-of-war', *Nat Commun*, 11: 6046.

Feng, R., M. Gruebele, and C. M. Davis. 2019. 'Quantifying protein dynamics and stability in a living organism', *Nat Commun*, 10: 1179.

Fowler, N. J., A. Sljoka, and M. P. Williamson. 2020. 'A method for validating the accuracy of NMR protein structures', *Nat Commun*, 11: 6321.

Fox, J. C., M. A. Thomas, A. F. Dishman, O. Larsen, T. Nakayama, O. Yoshie, M. M. Rosenkilde, and B. F. Volkman. 2019. 'Structure-function guided modeling of chemokine-GPCR specificity for the chemokine XCL1 and its receptor XCR1', *Sci Signal*, 12.

Froning, K., J. Maguire, A. Sereno, F. Huang, S. Chang, K. Weichert, A. J. Frommelt, J. Dong, X. Wu, H. Austin, E. M. Conner, J. R. Fitchett, A. R. Heng, D. Balasubramaniam, M. T. Hilgers, B. Kuhlman,

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

and S. J. Demarest. 2020. 'Computational stabilization of T cell receptors allows pairing with antibodies to form bispecifics', *Nat Commun*, 11: 2330.

Gabryelczyk, B., H. Cai, X. Shi, Y. Sun, P. J. M. Swinkels, S. Salentinig, K. Pervushin, and A. Miserez. 2019. 'Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides', *Nat Commun*, 10: 5465.

Gaglia, G., R. Rashid, C. Yapp, G. N. Joshi, C. G. Li, S. L. Lindquist, K. A. Sarosiek, L. Whitesell, P. K. Sorger, and S. Santagata. 2020. 'HSF1 phase transition mediates stress adaptation and cell fate decisions', *Nat Cell Biol*, 22: 151-58.

Gainza, P., F. Sverrisson, F. Monti, E. Rodola, D. Boscaini, M. M. Bronstein, and B. E. Correia. 2020. 'Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning', *Nat Methods*, 17: 184-92.

Gelenter, M. D., K. J. Smith, S. Y. Liao, V. S. Mandala, A. J. Dregni, M. S. Lamm, Y. Tian, W. Xu, D. J. Pochan, T. J. Tucker, Y. Su, and M. Hong. 2019. 'The peptide hormone glucagon forms amyloid fibrils with two coexisting beta-strand conformations', *Nat Struct Mol Biol*, 26: 592-98.

Girstmair, H., F. Tippel, A. Lopez, K. Tych, F. Stein, P. Haberkant, P. W. N. Schmid, D. Helm, M. Rief, M. Sattler, and J. Buchner. 2019. 'The Hsp90 isoforms from S. cerevisiae differ in structure, function and client range', *Nat Commun*, 10: 3626.

Gladrow, J., M. Ribezzi-Crivellari, F. Ritort, and U. F. Keyser. 2019. 'Experimental evidence of symmetry breaking of transition-path times', *Nat Commun*, 10: 55.

Glasgow, A. A., Y. M. Huang, D. J. Mandell, M. Thompson, R. Ritterson, A. L. Loshbaugh, J. Pellegrino, C. Krivacic, R. A. Pache, K. A. Barlow, N. Ollikainen, D. Jeon, M. J. S. Kelly, J. S. Fraser, and T. Kortemme. 2019. 'Computational design of a modular protein sense-response system', *Science*, 366: 1024-28.

Guo, Y., M. Vucelja, and A. Amir. 2019. 'Stochastic tunneling across fitness valleys can give rise to a logarithmic long-term fitness trajectory', *Sci Adv*, 5: eaav3842.

Hannich, J. T., A. G. Haribowo, S. Gentina, M. Paillard, L. Gomez, B. Pillot, H. Thibault, D. Abegg, N. Guex, A. Zumbuehl, A. Adibekian, M. Ovize, J. C. Martinou, and H. Riezman. 2019. '1-Deoxydihydroceramide causes anoxic death by impairing chaperonin-mediated protein folding', *Nat Metab*, 1: 996-1008.

Hetz, C., K. Zhang, and R. J. Kaufman. 2020. 'Mechanisms, regulation and functions of the unfolded protein response', *Nat Rev Mol Cell Biol*, 21: 421-38.

Heyne, H. O., D. Baez-Nieto, S. Iqbal, D. S. Palmer, A. Brunklaus, P. May, Collaborative Epi, K. M. Johannesen, S. Lauxmann, J. R. Lemke, R. S. Moller, E. Perez-Palma, U. I. Scholl, S. Syrbe, H. Lerche, D. Lal, A. J. Campbell, H. R. Wang, J. Pan, and M. J. Daly. 2020. 'Predicting functional effects of missense variants in voltage-gated sodium and calcium channels', *Sci Transl Med*, 12.

Hipp, M. S., P. Kasturi, and F. U. Hartl. 2019. 'The proteostasis network and its decline in ageing', *Nat Rev Mol Cell Biol*, 20: 421-35.

Ho, C. T., T. Grousl, O. Shatz, A. Jawed, C. Ruger-Herreros, M. Semmelink, R. Zahn, K. Richter, B. Bukau, and A. Mogk. 2019. 'Cellular sequestrases maintain basal Hsp70 capacity ensuring balanced proteostasis', *Nat Commun*, 10: 4851.

Huang, J. X., G. Lee, K. E. Cavanaugh, J. W. Chang, M. L. Gardel, and R. E. Moellering. 2019. 'High throughput discovery of functional protein modifications by Hotspot Thermal Profiling', *Nat Methods*, 16: 894-901.

Imamoglu, R., D. Balchin, M. Hayer-Hartl, and F. U. Hartl. 2020. 'Bacterial Hsp70 resolves misfolded states and accelerates productive folding of a multi-domain protein', *Nat Commun*, 11: 365.

Jalan, A. A., D. Sammon, J. D. Hartgerink, P. Brear, K. Stott, S. W. Hamaia, E. J. Hunter, D. R. Walker, B. Leitinger, and R. W. Farndale. 2020. 'Chain alignment of collagen I deciphered using computationally designed heterotrimers', *Nat Chem Biol*, 16: 423-29.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Jarzab, A., N. Kurzawa, T. Hopf, M. Moerch, J. Zecha, N. Leijten, Y. Bian, E. Musiol, M. Maschberger, G. Stoehr, I. Becher, C. Daly, P. Samaras, J. Mergner, B. Spanier, A. Angelov, T. Werner, M. Bantscheff, M. Wilhelm, M. Klingenspor, S. Lemeer, W. Liebl, H. Hahne, M. M. Savitski, and B. Kuster. 2020. 'Meltome atlas-thermal proteome stability across the tree of life', *Nat Methods*, 17: 495-503.

Jiang, Y., P. Rossi, and C. G. Kalodimos. 2019. 'Structural basis for client recognition and activity of Hsp40 chaperones', *Science*, 365: 1313-19.

Joazeiro, C. A. P. 2019. 'Mechanisms and functions of ribosome-associated protein quality control', *Nat Rev Mol Cell Biol*, 20: 368-83.

Johansson, L. C., B. Stauch, J. D. McCorvy, G. W. Han, N. Patel, X. P. Huang, A. Batyuk, C. Gati, S. T. Slocum, C. Li, J. M. Grandner, S. Hao, R. H. J. Olsen, A. R. Tribo, S. Zaare, L. Zhu, N. A. Zatsepin, U. Weierstall, S. Yous, R. C. Stevens, W. Liu, B. L. Roth, V. Katritch, and V. Cherezov. 2019. 'XFEL structures of the human MT2 melatonin receptor reveal the basis of subtype selectivity', *Nature*, 569: 289-92.

Joseph, B., E. A. Jaumann, A. Sikora, K. Barth, T. F. Prisner, and D. S. Cafiso. 2019. 'In situ observation of conformational dynamics and protein ligand-substrate interactions in outer-membrane proteins with DEER/PELDOR spectroscopy', *Nat Protoc*, 14: 2344-69.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. 2021. 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596: 583-89.

Kabayama, H., M. Takeuchi, N. Tokushige, S. I. Muramatsu, M. Kabayama, M. Fukuda, Y. Yamada, and K. Mikoshiba. 2020. 'An ultra-stable cytoplasmic antibody engineered for in vivo applications', *Nat Commun*, 11: 336.

Kalxdorf, M., I. Gunthner, I. Becher, N. Kurzawa, S. Knecht, M. M. Savitski, H. C. Eberl, and M. Bantscheff. 2021. 'Cell surface thermal proteome profiling tracks perturbations and drug targets on the plasma membrane', *Nat Methods*, 18: 84-91.

Khoury, L. R., and I. Popa. 2019. 'Chemical unfolding of protein domains induces shape change in programmed protein hydrogels', *Nat Commun*, 10: 5439.

Khoury, L. R., M. Slawinski, D. R. Collison, and I. Popa. 2020. 'Cation-induced shape programming and morphing in protein-based hydrogels', *Sci Adv*, 6: eaba6112.

Klaips, C. L., M. H. M. Gropp, M. S. Hipp, and F. U. Hartl. 2020. 'Sis1 potentiates the stress response to protein aggregation and elevated temperature', *Nat Commun*, 11: 6271.

Koepnick, B., J. Flatten, T. Husain, A. Ford, D. A. Silva, M. J. Bick, A. Bauer, G. Liu, Y. Ishida, A. Boykov, R. D. Estep, S. Kleinfelter, T. Norgard-Solano, L. Wei, F. Players, G. T. Montelione, F. DiMaio, Z. Popovic, F. Khatib, S. Cooper, and D. Baker. 2019. 'De novo protein design by citizen scientists', *Nature*, 570: 390-94.

Kohn, B., and M. Kovermann. 2020. 'All atom insights into the impact of crowded environments on protein stability by NMR spectroscopy', *Nat Commun*, 11: 5760.

Kopp, M. C., N. Larburu, V. Durairaj, C. J. Adams, and M. M. U. Ali. 2019. 'UPR proteins IRE1 and PERK switch BiP from chaperone to ER stress sensor', *Nat Struct Mol Biol*, 26: 1053-62.

Kuhlman, B., and P. Bradley. 2019. 'Advances in protein structure prediction and design', *Nat Rev Mol Cell Biol*, 20: 681-97.

Kumar, N., C. P. Arthur, C. Ciferri, and M. L. Matsumoto. 2020. 'Structure of the secretory immunoglobulin A core', *Science*, 367: 1008-14.

Langan, R. A., S. E. Boyken, A. H. Ng, J. A. Samson, G. Dods, A. M. Westbrook, T. H. Nguyen, M. J. Lajoie, Z. Chen, S. Berger, V. K. Mulligan, J. E. Dueber, W. R. P. Novak, H. El-Samad, and D. Baker. 2019. 'De novo design of bioactive protein switches', *Nature*, 572: 205-10.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Lee, J., Z. Liu, P. H. Suzuki, J. F. Ahrens, S. Lai, X. Lu, S. Guan, and F. St-Pierre. 2020. 'Versatile phenotype-activated cell sorting', *Sci Adv*, 6.

Leopold, A. V., K. G. Chernov, A. A. Shemetov, and V. V. Verkhusha. 2019. 'Neurotrophin receptor tyrosine kinases regulated with near-infrared light', *Nat Commun*, 10: 1129.

Li, G., Y. Hu, Zrimec Jan, H. Luo, H. Wang, A. Zelezniak, B. Ji, and J. Nielsen. 2021. 'Bayesian genome scale modelling identifies thermal determinants of yeast metabolism', *Nat Commun*, 12: 190.

Li, Y., K. Li, X. Wang, M. Cui, P. Ge, J. Zhang, F. Qiu, and C. Zhong. 2020. 'Conformable self-assembling amyloid protein coatings with genetically programmable functionality', *Sci Adv*, 6: eaba1425.

Li, Y., D. Sun, Z. Ma, K. Yamaguchi, L. Wang, S. Zhong, X. Yan, B. Shang, Y. Nagashima, H. Koiwa, J. Han, Q. Xie, M. Zhou, Z. Wang, and X. Zhang. 2020. 'Degradation of SERRATE via ubiquitin-independent 20S proteasome to survey RNA metabolism', *Nat Plants*, 6: 970-82.

Liberta, F., S. Loerch, M. Rennegarbe, A. Schierhorn, P. Westermark, G. T. Westermark, B. P. C. Hazenberg, N. Grigorieff, M. Fandrich, and M. Schmidt. 2019. 'Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids', *Nat Commun*, 10: 1104.

Liess, A. K. L., A. Kucerova, K. Schweimer, D. Schlesinger, O. Dybkov, H. Urlaub, J. Mansfeld, and S. Lorenz. 2020. 'Dimerization regulates the human APC/C-associated ubiquitin-conjugating enzyme UBE2S', *Sci Signal*, 13.

Lim, S., H. Y. Cho, D. G. Kim, Y. Roh, S. Y. Son, A. U. Mushtaq, M. Kim, D. Bhattarai, A. Sivaraman, Y. Lee, J. Lee, W. S. Yang, H. K. Kim, M. H. Kim, K. Lee, Y. H. Jeon, and S. Kim. 2020. 'Targeting the interaction of AIMP2-DX2 with HSP70 suppresses cancer development', *Nat Chem Biol*, 16: 31-41.

Lin, X., M. Li, N. Wang, Y. Wu, Z. Luo, S. Guo, G. W. Han, S. Li, Y. Yue, X. Wei, X. Xie, Y. Chen, S. Zhao, J. Wu, M. Lei, and F. Xu. 2020. 'Structural basis of ligand recognition and self-activation of orphan GPR52', *Nature*, 579: 152-57.

Liu, J., D. Guan, M. Dong, J. Yang, H. Wei, Q. Liang, L. Song, L. Xu, J. Bai, C. Liu, J. Mao, Q. Zhang, J. Zhou, X. Wu, M. Wang, and Y. S. Cong. 2020. 'UFMylation maintains tumour suppressor p53 stability by antagonizing its ubiquitination', *Nat Cell Biol*, 22: 1056-63.

Liu, X. R., D. L. Rempel, and M. L. Gross. 2020. 'Protein higher-order-structure determination by fast photochemical oxidation of proteins and mass spectrometry analysis', *Nat Protoc*, 15: 3942-70.

Liu, Z., H. Liu, A. M. Vera, R. C. Bernardi, P. Tinnefeld, and M. A. Nash. 2020. 'High force catch bond mechanism of bacterial adhesion in the human gut', *Nat Commun*, 11: 4321.

Louros, N., G. Orlando, M. De Vleeschouwer, F. Rousseau, and J. Schymkowitz. 2020. 'Structure-based machine-guided mapping of amyloid sequence space reveals uncharted sequence clusters with higher solubilities', *Nat Commun*, 11: 3314.

Mann, J. L., C. L. Maikawa, A. A. A. Smith, A. K. Grosskopf, S. W. Baker, G. A. Roth, C. M. Meis, E. C. Gale, C. S. Liong, S. Correa, D. Chan, L. M. Stapleton, A. C. Yu, B. Muir, S. Howard, A. Postma, and E. A. Appel. 2020. 'An ultrafast insulin formulation enabled by high-throughput screening of engineered polymeric excipients', *Sci Transl Med*, 12.

Marchand, A., M. F. Czar, E. N. Eggel, J. Kaeslin, and R. Zenobi. 2020. 'Studying biomolecular folding and binding using temperature-jump mass spectrometry', *Nat Commun*, 11: 566.

Martens, C., M. Shekhar, A. M. Lau, E. Tajkhorshid, and A. Politis. 2019. 'Integrating hydrogen-deuterium exchange mass spectrometry with molecular dynamics simulations to probe lipid-modulated conformational changes in membrane proteins', *Nat Protoc*, 14: 3183-204.

Mas, G., B. M. Burmann, T. Sharpe, B. Claudi, D. Bumann, and S. Hiller. 2020. 'Regulation of chaperone function by coupled folding and oligomerization', *Sci Adv*, 6.

Mateus, A., J. Hevler, J. Bobonis, N. Kurzawa, M. Shah, K. Mitosch, C. V. Goemans, D. Helm, F. Stein, A. Typas, and M. M. Savitski. 2020. 'The functional proteome landscape of Escherichia coli', *Nature*, 588: 473-78.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Michiels, E., K. Roose, R. Gallardo, L. Khodaparast, L. Khodaparast, R. van der Kant, M. Siemons, B. Houben, M. Ramakers, H. Wilkinson, P. Guerreiro, N. Louros, S. J. F. Kaptein, L. I. Ibanez, A. Smet, P. Baatsen, S. Liu, I. Vorberg, G. Bormans, J. Neyts, X. Saelens, F. Rousseau, and J. Schymkowitz. 2020. 'Reverse engineering synthetic antiviral amyloids', *Nat Commun*, 11: 2832.

Mogk, A., and B. Bukau. 2017. 'Role of sHsps in organizing cytosolic protein aggregation and disaggregation', *Cell Stress Chaperones*, 22: 493-502.

Mueller, F., A. Friese, C. Pathe, R. C. da Silva, K. B. Rodriguez, A. Musacchio, and T. Bange. 2021. 'Overlap of NatA and IAP substrates implicates N-terminal acetylation in protein stabilization', *Sci Adv*, 7.

Muller, J. B., P. E. Geyer, A. R. Colaco, P. V. Treit, M. T. Strauss, M. Oroshi, S. Doll, S. Virreira Winter, J. M. Bader, N. Kohler, F. Theis, A. Santos, and M. Mann. 2020. 'The proteome landscape of the kingdoms of life', *Nature*, 582: 592-96.

Nandwani, N., P. Surana, H. Negi, N. M. Mascarenhas, J. B. Udgaonkar, R. Das, and S. Gosavi. 2019. 'A five-residue motif for the design of domain swapping in proteins', *Nat Commun*, 10: 452.

Ni, T., F. Jiao, X. Yu, S. Aden, L. Ginger, S. I. Williams, F. Bai, V. Prazak, D. Karia, P. Stansfeld, P. Zhang, G. Munson, G. Anderluh, S. Scheuring, and R. J. C. Gilbert. 2020. 'Structure and mechanism of bactericidal mammalian perforin-2, an ancient agent of innate immunity', *Sci Adv*, 6: eaax8286.

Nirmalraj, P. N., J. List, S. Battacharya, G. Howe, L. Xu, D. Thompson, and M. Mayer. 2020. 'Complete aggregation pathway of amyloid beta (1-40) and (1-42) resolved on an atomically clean interface', *Sci Adv*, 6: eaaz6014.

Noe, F., S. Olsson, J. Kohler, and H. Wu. 2019. 'Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning', *Science*, 365.

Okumura, M., K. Noi, S. Kanemura, M. Kinoshita, T. Saio, Y. Inoue, T. Hikima, S. Akiyama, T. Ogura, and K. Inaba. 2019. 'Dynamic assembly of protein disulfide isomerase in catalysis of oxidative folding', *Nat Chem Biol*, 15: 499-509.

Oliinyk, O. S., A. A. Shemetov, S. Pletnev, D. M. Shcherbakova, and V. V. Verkhusha. 2019. 'Smallest near-infrared fluorescent protein evolved from cyanobacteriochrome as versatile tag for spectral multiplexing', *Nat Commun*, 10: 279.

Orlando, G., D. Raimondi, and F. Vranken W. 2019. 'Auto-encoding NMR chemical shifts from their native vector space to a residue-level biophysical index', *Nat Commun*, 10: 2511.

Ou, X., B. Xue, Y. Lao, Y. Wutthinitikornkit, R. Tian, A. Zou, L. Yang, W. Wang, Y. Cao, and J. Li. 2020. 'Structure and sequence features of mussel adhesive protein lead to its salt-tolerant adhesion ability', *Sci Adv*, 6.

Pan, X., M. C. Thompson, Y. Zhang, L. Liu, J. S. Fraser, M. J. S. Kelly, and T. Kortemme. 2020. 'Expanding the space of protein geometries by computational design of de novo fold families', *Science*, 369: 1132-36.

Pankow, S., C. Bamberger, and J. R. Yates, 3rd. 2019. 'A posttranslational modification code for CFTR maturation is altered in cystic fibrosis', *Sci Signal*, 12.

Peacock, A. 2020. 'Can proteins be truly designed sans function?', *Science*, 369: 1166-67.

Pearce, P., F. G. Woodhouse, A. Forrow, A. Kelly, H. Kusumaatmaja, and J. Dunkel. 2019. 'Learning dynamical information from static protein and sequencing data', *Nat Commun*, 10: 5368.

Penn, W. D., A. G. McKee, C. P. Kuntz, H. Woods, V. Nash, T. C. Gruenhagen, F. J. Roushar, M. Chandak, C. Hemmerich, D. B. Rusch, J. Meiler, and J. P. Schlebach. 2020. 'Probing biophysical sequence constraints within the transmembrane domains of rhodopsin by deep mutational scanning', *Sci Adv*, 6: eaay7505.

Perrin, J., T. Werner, N. Kurzawa, A. Rutkowska, D. D. Childs, M. Kalxdorf, D. Poeckel, E. Stonehouse, K. Strohmer, B. Heller, D. W. Thomson, J. Krause, I. Becher, H. C. Eberl, J. Vappiani, D. C. Sevin, C. E. Rau, H. Franken, W. Huber, M. Faelth-Savitski, M. M. Savitski, M. Bantscheff, and G. Bergamini. 2020. 'Identifying drug targets in tissues and whole blood with thermal-shift profiling', *Nat Biotechnol*, 38: 303-08.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Pillai, A. S., S. A. Chandler, Y. Liu, A. V. Signore, C. R. Cortez-Romero, J. L. P. Benesch, A. Laganowsky, J. F. Storz, G. K. A. Hochberg, and J. W. Thornton. 2020. 'Origin of complexity in haemoglobin evolution', *Nature*, 581: 480-85.

Poelwijk, F. J., M. Socolich, and R. Ranganathan. 2019. 'Learning the pattern of epistasis linking genotype and phenotype in a protein', *Nat Commun*, 10: 4213.

Polizzi, N. F., and W. F. DeGrado. 2020. 'A defined structural unit enables de novo design of small-molecule-binding proteins', *Science*, 369: 1227-33.

Quadeer, A. A., R. H. Y. Louie, and M. R. McKay. 2019. 'Identifying immunologically-vulnerable regions of the HCV E2 glycoprotein and broadly neutralizing antibodies that target them', *Nat Commun*, 10: 2073.

Radamaker, L., Y. H. Lin, K. Annamalai, S. Huhn, U. Hegenbart, S. O. Schonland, G. Fritz, M. Schmidt, and M. Fandrich. 2019. 'Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis', *Nat Commun*, 10: 1103.

Rizzolo, K., A. Y. H. Yu, A. Ologbenla, S. R. Kim, H. Zhu, K. Ishimori, G. Thibault, E. Leung, Y. W. Zhang, M. Teng, M. Haniszewski, N. Miah, S. Phanse, Z. Minic, S. Lee, J. D. Caballero, M. Babu, F. T. F. Tsai, T. Saio, and W. A. Houry. 2021. 'Functional cooperativity between the trigger factor chaperone and the ClpXP proteolytic complex', *Nat Commun*, 12: 281.

Roder, C., T. Kupreichyk, L. Gremer, L. U. Schafer, K. R. Pothula, R. B. G. Ravelli, D. Willbold, W. Hoyer, and G. F. Schroder. 2020. 'Cryo-EM structure of islet amyloid polypeptide fibrils reveals similarities with amyloid-beta fibrils', *Nat Struct Mol Biol*, 27: 660-67.

Roder, C., N. Vettore, L. N. Mangels, L. Gremer, R. B. G. Ravelli, D. Willbold, W. Hoyer, A. K. Buell, and G. F. Schroder. 2019. 'Atomic structure of PI3-kinase SH3 amyloid fibrils by cryo-electron microscopy', *Nat Commun*, 10: 3754.

Rollins, N. J., K. P. Brock, F. J. Poelwijk, M. A. Stiffler, N. P. Gauthier, C. Sander, and D. S. Marks. 2019. 'Inferring protein 3D structure from deep mutation scans', *Nat Genet*, 51: 1170-76.

Sang, P., Z. Zhou, Y. Shi, C. Lee, Z. Amso, D. Huang, T. Odom, V. T. B. Nguyen-Tran, W. Shen, and J. Cai. 2020. 'The activity of sulfono-gamma-AApeptide helical foldamers that mimic GLP-1', *Sci Adv*, 6: eaaz4988.

Schapira, M., M. F. Calabrese, A. N. Bullock, and C. M. Crews. 2019. 'Targeted protein degradation: expanding the toolbox', *Nat Rev Drug Discov*, 18: 949-63.

Schmidt, M., S. Wiese, V. Adak, J. Engler, S. Agarwal, G. Fritz, P. Westermark, M. Zacharias, and M. Fandrich. 2019. 'Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis', *Nat Commun*, 10: 5008.

Schmiedel, J. M., and B. Lehner. 2019. 'Determining protein structures using deep mutagenesis', *Nat Genet*, 51: 1177-86.

Schober, R., D. Bonhenry, V. Lunz, J. Zhu, A. Krizova, I. Frischauf, M. Fahrner, M. Zhang, L. Waldherr, T. Schmidt, I. Derler, P. B. Stathopulos, C. Romanin, R. H. Ettrich, and R. Schindl. 2019. 'Sequential activation of STIM1 links Ca(2+) with luminal domain unfolding', *Sci Signal*, 12.

Schulte, L., J. Mao, J. Reitz, S. Sreeramulu, D. Kudlinzki, V. V. Hodirnau, J. Meier-Credo, K. Saxena, F. Buhr, J. D. Langer, M. Blackledge, A. S. Frangakis, C. Glaubitz, and H. Schwalbe. 2020. 'Cysteine oxidation and disulfide formation in the ribosomal exit tunnel', *Nat Commun*, 11: 5569.

Sedlak, S. M., L. C. Schendel, H. E. Gaub, and R. C. Bernardi. 2020. 'Streptavidin/biotin: Tethering geometry defines unbinding mechanics', *Sci Adv*, 6: eaay5999.

Serdiuk, T., A. Steudle, S. A. Mari, S. Manioglu, H. R. Kaback, A. Kuhn, and D. J. Muller. 2019. 'Insertion and folding pathways of single membrane proteins guided by translocases and insertases', *Sci Adv*, 5: eaau6824.

Sesterhenn, F., C. Yang, J. Bonet, J. T. Cramer, X. Wen, Y. Wang, C. I. Chiang, L. A. Abriata, I. Kucharska, G. Castoro, S. S. Vollers, M. Galloux, E. Dheilly, S. Rosset, P. Corthesy, S. Georgeon, M. Villard, C. A. Richard, D. Descamps, T. Delgado, E. Oricchio, M. A. Rameix-Welti, V. Mas, S. Ervin, J. F. Eleouet, S. Riffault, J. T. Bates, J. P. Julien, Y. Li, T. Jardetzky, T. Krey, and B. E. Correia. 2020.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

'De novo protein design enables the precise induction of RSV-neutralizing antibodies', *Science*, 368.

Shao, W., T. Guo, N. C. Toussaint, P. Xue, U. Wagner, L. Li, K. Charmpi, Y. Zhu, J. Wu, M. Buljan, R. Sun, D. Rutishauser, T. Hermanns, C. D. Fankhauser, C. Poyet, J. Ljubicic, N. Rupp, J. H. Ruschoff, Q. Zhong, A. Beyer, J. Ji, B. C. Collins, Y. Liu, G. Ratsch, P. J. Wild, and R. Aebersold. 2019. 'Comparative analysis of mRNA and protein degradation in prostate tissues indicates high stability of proteins', *Nat Commun*, 10: 2524.

Shin, C. S., S. Meng, S. D. Garbis, A. Moradian, R. W. Taylor, M. J. Sweredoski, B. Lomenick, and D. C. Chan. 2021. 'LONP1 and mtHSP70 cooperate to promote mitochondrial protein folding', *Nat Commun*, 12: 265.

Shishido, H., J. S. Yoon, Z. Yang, and W. R. Skach. 2020. 'CFTR trafficking mutations disrupt cotranslational protein folding by targeting biosynthetic intermediates', *Nat Commun*, 11: 4258.

Shohei Eda, Igor Nasibullin, Kenward Vong, Norio Kudo, Minoru Yoshida, Almira Kurbangalieva, and Katsunori Tanaka. 2019. 'Biocompatibility and therapeutic potential of glycosylated albumin artificial metalloenzymes', *Nature Catalysis*, 2: 780-92.

Sigoillot, M., M. Overtus, M. Grodecka, D. Scholl, A. Garcia-Pino, T. Laeremans, L. He, E. Pardon, E. Hildebrandt, I. Urbatsch, J. Steyaert, J. R. Riordan, and C. Govaerts. 2019. 'Domain-interface dynamics of CFTR revealed by stabilizing nanobodies', *Nat Commun*, 10: 2636.

Smole, U., N. Gour, J. Phelan, G. Hofer, C. Kohler, B. Kratzer, P. A. Tauber, X. Xiao, N. Yao, J. Dvorak, L. Caraballo, L. Puerta, S. Rosskopf, J. Chakir, E. Malle, A. P. Lane, W. F. Pickl, S. Lajoie, and M. Wills-Karp. 2020. 'Serum amyloid A is a soluble pattern recognition receptor that drives type 2 immunity', *Nat Immunol*, 21: 756-65.

Socan, J., M. Purg, and J. Aqvist. 2020. 'Computer simulations explain the anomalous temperature optimum in a cold-adapted enzyme', *Nat Commun*, 11: 2644.

Song, K. H., S. J. Oh, S. Kim, H. Cho, H. J. Lee, J. S. Song, J. Y. Chung, E. Cho, J. Lee, S. Jeon, C. Yee, K. M. Lee, S. M. Hewitt, J. H. Kim, S. R. Woo, and T. W. Kim. 2020. 'HSP90A inhibition promotes anti-tumor immunity by reversing multi-modal resistance and stem-like property of immune-refractory tumors', *Nat Commun*, 11: 562.

Spohn, R., L. Daruka, V. Lazar, A. Martins, F. Vidovics, G. Grezal, O. Mehi, B. Kintses, M. Szamel, P. K. Jangir, B. Csorgo, A. Gyorkei, Z. Bodi, A. Farago, L. Bodai, I. Foldesi, D. Kata, G. Maroti, B. Pap, R. Wirth, B. Papp, and C. Pal. 2019. 'Integrated evolutionary analysis reveals antimicrobial peptides with limited resistance', *Nat Commun*, 10: 4538.

Srivastava, D., L. Gakhar, and N. O. Artemyev. 2019. 'Structural underpinnings of Ric8A function as a G-protein alpha-subunit chaperone and guanine-nucleotide exchange factor', *Nat Commun*, 10: 3084.

Stohr, M., and A. Tkatchenko. 2019. 'Quantum mechanics of proteins in explicit water: The role of plasmon-like solute-solvent interactions', *Sci Adv*, 5: eaax0024.

Stone, N. P., G. Demo, E. Agnello, and B. A. Kelch. 2019. 'Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure', *Nat Commun*, 10: 4471.

Sun, H., and B. Marelli. 2020. 'Polypeptide templating for designer hierarchical materials', *Nat Commun*, 11: 351.

Sun, Y., K. Zhao, W. Xia, G. Feng, J. Gu, Y. Ma, X. Gui, X. Zhang, Y. Fang, B. Sun, R. Wang, C. Liu, and D. Li. 2020. 'The nuclear localization sequence mediates hnRNPA1 amyloid fibril formation revealed by cryoEM structure', *Nat Commun*, 11: 6349.

Swuec, P., F. Lavatelli, M. Tasaki, C. Paissoni, P. Rognoni, M. Maritan, F. Brambilla, P. Milani, P. Mauri, C. Camilloni, G. Palladini, G. Merlini, S. Ricagno, and M. Bolognesi. 2019. 'Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient', *Nat Commun*, 10: 1269.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

Tang, Z., K. H. Su, M. Xu, and C. Dai. 2020. 'HSF1 physically neutralizes amyloid oligomers to empower overgrowth and bestow neuroprotection', *Sci Adv*, 6.

Tapia-Rojo, R., A. Alonso-Caballero, and J. M. Fernandez. 2020. 'Direct observation of a coil-to-helix contraction triggered by vinculin binding to talin', *Sci Adv*, 6: eaaz4707.

Tong Wang, Yanhua Qiao, Wenze Ding, Wenzhi Mao, Yaoqi Zhou, and Haipeng Gong. 2019. 'Improved fragment sampling for ab initio protein structure prediction using deep neural networks', *Nature Machine Intelligence*, 1: 347–55.

Toyama, Y., K. Kontani, T. Katada, and I. Shimada. 2019. 'Decreased conformational stability in the oncogenic N92I mutant of Ras-related C3 botulinum toxin substrate 1', *Sci Adv*, 5: eaax1595.

Tsurusawa, H., S. Arai, and H. Tanaka. 2020. 'A unique route of colloidal phase separation yields stress-free gels', *Sci Adv*, 6.

Upmeier zu Belzen, Julius, Thore Bürgel, Stefan Holderbach, Felix Bubeck, Lukas Adam, Catharina Gandor, Marita Klein, Jan Mathony, Pauline Pfuderer, Lukas Platz, Moritz Przybilla, Max Schwendemann, Daniel Heid, Mareike Daniela Hoffmann, Michael Jendrusch, Carolin Schmelas, Max Waldhauer, Irina Lehmann, Dominik Niopek, and Roland Eils. 2019. 'Leveraging implicit knowledge in neural networks for functional dissection and engineering of proteins', *Nature Machine Intelligence*, 1: 225-35.

Vimer, S., G. Ben-Nissan, and M. Sharon. 2020. 'Direct characterization of overproduced proteins by native mass spectrometry', *Nat Protoc*, 15: 236-65.

Wang, J., A. Jain, L. R. McDonald, C. Gambogi, A. L. Lee, and N. V. Dokholyan. 2020. 'Mapping allosteric communications within individual proteins', *Nat Commun*, 11: 3862.

Wang, L., L. Zhang, L. Li, J. Jiang, Z. Zheng, J. Shang, C. Wang, W. Chen, Q. Bao, X. Xu, Z. Jiang, J. Zhang, and Q. You. 2019. 'Small-molecule inhibitor targeting the Hsp90-Cdc37 protein-protein interaction in colorectal cancer', *Sci Adv*, 5: eaax2277.

Wang, M., Z. Cang, and G. W. Wei. 2020. 'A topology-based network tree for the prediction of protein-protein binding affinity changes following mutation', *Nat Mach Intell*, 2: 116-23.

Wenzhi Mao, Wenze Ding, Yaoguang Xing, and Haipeng Gong. 2020. 'AmoebaContact and GDFold as a pipeline for rapid de novo protein structure prediction', *Nature machine inteligence*: 25-33.

Williams, T. A., C. J. Cox, P. G. Foster, G. J. Szollosi, and T. M. Embley. 2020. 'Phylogenomics provides robust support for a two-domains tree of life', *Nat Ecol Evol*, 4: 138-47.

Wu, G. Z., E. H. Meyer, A. S. Richter, M. Schuster, Q. Ling, M. A. Schottler, D. Walther, R. Zoschke, B. Grimm, R. P. Jarvis, and R. Bock. 2019. 'Control of retrograde signalling by protein import and cytosolic folding stress', *Nat Plants*, 5: 525-38.

Wu, K., F. Stull, C. Lee, and J. C. A. Bardwell. 2019. 'Protein folding while chaperone bound is dependent on weak interactions', *Nat Commun*, 10: 4833.

Wu, X., M. Siggel, S. Ovchinnikov, W. Mi, V. Svetlov, E. Nudler, M. Liao, G. Hummer, and T. A. Rapoport. 2020. 'Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex', *Science*, 368.

Xhani, S., S. Lee, H. M. Kim, S. Wang, S. Esaki, V. L. T. Ha, M. Khanezarrin, G. L. Fernandez, A. V. Albrecht, J. M. Aramini, M. W. Germann, and G. M. K. Poon. 2020. 'Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1', *Sci Adv*, 6: eaay3178.

Xiao, P., D. Bolton, R. A. Munro, L. S. Brown, and V. Ladizhansky. 2019. 'Solid-state NMR spectroscopy based atomistic view of a membrane protein unfolding pathway', *Nat Commun*, 10: 3867.

Xiong, X., K. Qu, K. A. Ciazynska, M. Hosmillo, A. P. Carter, S. Ebrahimi, Z. Ke, S. H. W. Scheres, L. Bergamaschi, G. L. Grice, Y. Zhang, Citiid-Nihr Covid- BioResource Collaboration, J. A. Nathan, S. Baker, L. C. James, H. E. Baxendale, I. Goodfellow, R. Doffinger, and J. A. G. Briggs. 2020. 'A thermostable, closed SARS-CoV-2 spike protein trimer', *Nat Struct Mol Biol*, 27: 934-41.

Xu, C., P. Lu, T. M. Gamal El-Din, X. Y. Pei, M. C. Johnson, A. Uyeda, M. J. Bick, Q. Xu, D. Jiang, H. Bai, G. Reggiano, Y. Hsia, T. J. Brunette, J. Dou, D. Ma, E. M. Lynch, S. E. Boyken, P. S. Huang, L. Stewart,

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333

F. DiMaio, J. M. Kollman, B. F. Luisi, T. Matsuura, W. A. Catterall, and D. Baker. 2020. 'Computational design of transmembrane pores', *Nature*, 585: 129-34.

Yan, X., X. Wang, Y. Li, M. Zhou, Y. Li, L. Song, W. Mi, J. Min, and C. Dong. 2021. 'Molecular basis for ubiquitin ligase CRL2(FEM1C)-mediated recognition of C-degron', *Nat Chem Biol*, 17: 263-71.

Yang, C., F. Sesterhenn, J. Bonet, E. A. van Aalen, L. Scheller, L. A. Abriata, J. T. Cramer, X. Wen, S. Rosset, S. Georgeon, T. Jardetzky, T. Krey, M. Fussenegger, M. Merkx, and B. E. Correia. 2021. 'Bottom-up de novo design of functional proteins with complex structural features', *Nat Chem Biol*, 17: 492-500.

Yao, X., C. Chen, Y. Wang, S. Dong, Y. J. Liu, Y. Li, Z. Cui, W. Gong, S. Perrett, L. Yao, R. Lamed, E. A. Bayer, Q. Cui, and Y. Feng. 2020. 'Discovery and mechanism of a pH-dependent dual-binding-site switch in the interaction of a pair of protein modules', *Sci Adv*, 6.

Yee, A. W., M. Aldeghi, M. P. Blakeley, A. Ostermann, P. J. Mas, M. Moulin, D. de Sanctis, M. W. Bowler, C. Mueller-Dieckmann, E. P. Mitchell, M. Haertlein, B. L. de Groot, E. Boeri Erba, and V. T. Forsyth. 2019. 'A molecular mechanism for transthyretin amyloidogenesis', *Nat Commun*, 10: 925.

Yu, L., W. Zhang, W. Luo, R. L. Dupont, Y. Xu, Y. Wang, B. Tu, H. Xu, X. Wang, Q. Fang, Y. Yang, C. Wang, and C. Wang. 2020. 'Molecular recognition of human islet amyloid polypeptide assembly by selective oligomerization of thioflavin T', *Sci Adv*, 6: eabc1449.

Yu, M., Z. Zhao, Z. Chen, S. Le, and J. Yan. 2020. 'Modulating mechanical stability of heterodimerization between engineered orthogonal helical domains', *Nat Commun*, 11: 4476.

Zhang, Y., Y. Minagawa, H. Kizoe, K. Miyazaki, R. Iino, H. Ueno, K. V. Tabata, Y. Shimane, and H. Noji. 2019. 'Accurate high-throughput screening based on digital protein synthesis in a massively parallel femtoliter droplet array', *Sci Adv*, 5: eaav8185.

Zheng, J., N. Guo, and A. Wagner. 2020. 'Selection enhances protein evolvability by increasing mutational robustness and foldability', *Science*, 370.

Zimmermann, I., P. Egloff, C. A. J. Hutter, B. T. Kuhn, P. Brauer, S. Newstead, R. J. P. Dawson, E. R. Geertsma, and M. A. Seeger. 2020. 'Generation of synthetic nanobodies against delicate proteins', *Nat Protoc*, 15: 1707-41.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 952333